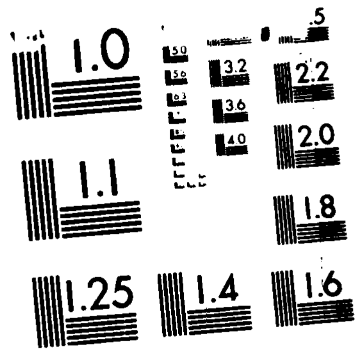
AD-A166 397 LOSS RATE ESTIMATION IN MARINE CORPS OFFICER MANPOWER MODELS(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA D D TUCKER SEP 85 1/2

UNCLASSIFIED F/G 5/9 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS

NAVAL POSTGRADUATE SCHOOL

Monterey, California

Supersedes AD-A162405 mc

A

THESIS

LOSS RATE ESTIMATION IN
MARINE CORPS OFFICER MANPOWER MODELS

by

Dewey Duane Tucker

September 1985

Thesis Advisor: Robert R. Read

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Loss Rate Estimation in Marine Corps Officer Manpower Models	5. TYPE OF REPORT & PERIOD COVERED Master's Thesis September, 1985
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dewey Duane Tucker	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943-5100	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943-5100	12. REPORT DATE September, 1985
	13. NUMBER OF PAGES 156
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Shrinkage Estimator, Attrition Rates, James Stein Estimators; Manpower Processes, Officer rate generator, Minimax

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Predicting officer attrition is a major difficulty in the Planning, Programming, and Budgeting Process. The Marine Corps has historically used aggregated estimates for attrition rates. The purpose of this paper is to demonstrate the application of the James-Stein and other shrinkage type parameter estimation schemes for the purpose of generating more stable manpower loss rates. It has been illustrated within this project that improvement can be attained by application of the James-Stein and Mimimax shrinkage

20. methods rather than the more natural Maximum Likelihood estimation process. The shrinkage schemes employed herein offer powerful and useful methods for generating attrition rates.

Approved for public release; distribution is unlimited.

Loss Rate Estimation in Marine Corps Officer Manpower Models

by

Dewey Duane Tucker
Major, United States Marine Corps
A.S., Ricks College, 1967
B.S., Oregon State University, 1971

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1985

Author: Dewey Duane Tucker

Approved by: Robert R. Read, Thesis Advisor

Paul R. Milch, Second Reader

Alan R. Washburn, Chairman,
Department of Operations Research

Kneale T. Marshall,
Dean of Information and Policy Sciences

ABSTRACT

Predicting officer attrition is a major difficulty in the Planning, Programming, and Budgeting Process. The Marine Corps has historically used aggregated estimates for attrition rates. The purpose of this paper is to demonstrate the application of the James-Stein and other shrinkage type parameter estimation schemes for the purpose of generating more stable manpower loss rates. It has been illustrated within this project that improvement can be attained by application of the James-Stein and Minimax shrinkage methods rather than the more natural Maximum Likelihood estimation process. The shrinkage schemes employed herein offer powerful and useful methods for generating attrition rates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. GENERAL

This paper is the result of a Headquarters Marine Corps request to explore the use of James-Stein and other shrinkage type parameter estimation schemes for the purpose of generating manpower loss rates. The very large number of cells within the USMC officer force structure leads to the condition that empirical attrition rates are unstable. This problem is compounded by the fact that the cell probabilities are small. Further difficulties are present because some of the inventory cells are empty for structural reasons while others are empty by chance. Therefore, the small inventory cells draw especial attention.

In order to deal with this phenomenon, variance stabilizing and symmetrizing transformations followed by versions of the James-Stein shrinkage technique were applied to selected aggregates of the USMC officer data. The application of these methods to the summary data provided by Navy Personnel Research and Development Center (NPRDC) has shown much promise. It has been illustrated within this project that improvement can be attained from application of these methods over other estimating methods.

Since considerable further work is needed to screen our choices, a further purpose of this thesis is to document a number of supporting materials, including a thorough discussion of the following:

1. Background of attrition and promotions for USMC officers.
2. Data types and the immediately available data base.

3. Identifying the plans (consumers) that utilize attrition rates.
4. Formulas and mathematical risk calculations relating to the James-Stein and empirical Bayes estimators.

Cross validation procedures were utilized to illustrate the improvement of shrinkage methods over the use of empirical rates.

B. BACKGROUND

The forecasts made by Manpower planning models are affected by three general factors: existing inventory, projected losses, and projected gains. In order to project the inventory into various future time periods it is necessary to use a realistic system of flow rates. Currently those rates are generated largely through arbitrary manual processes. The goal of this paper is to examine ways to develop objective and computerized loss rates.

There is an inherent confusion in terminology of losses and loss rates in that some are leavers from the Marine Corps and some are merely transitioning within the force structure. Flows of Officers from particular cells are characterized by Military Occupational Specialty (MOS), length of service (LOS), grade, and perhaps some other cross classifying characteristics. When the flows go from one cell to another within the Marine Corps, then the flow is referred to as a strength loss for the former cell. An officer moving from one cell to another cell does not mean a loss from the Marine Corps to the civilian labor market. Changes such as promotion to a higher pay grade or losses to a cell due to an officer increasing from one length of service to the next are other forms of transition within the force structure.

Losses from within the Marine Corps to the civilian labor market are the only losses treated in this thesis. They can be voluntary or involuntary. Voluntary losses occur when officers resign, retire, or are released by choice. Involuntary losses occur due to discharge, death, disability, release from active duty, and retirement. Retirement can be voluntary or involuntary.

Authors who have studied loss rates and how they apply to manpower planning are emphasizing the importance of understanding these rates in order to conduct proper forecasting. [Ref. 1] [Ref. 2]

A difficulty encountered in doing this project was that of obtaining sufficient data points with which to conduct cohort or census manpower data analysis [Ref. 1]. The reason for this difficulty was the state of the Marine Corps automated data reporting system prior to 1976. During the period 1970 to 1976, the Marine Corps was instituting its first automated personnel reporting system. Many problems were encountered during this period. Therefore, data from this period is generally unreliable. Data at the Defense Manpower Data Center is based on input from Headquarters Marine Corps and was found to have similar problems. A complicating factor during this period was the Vietnam conflict and the attritions generated by the conflict. These problems prevented an extended data base from which to draw from for the purpose of this paper. The Marine Corps manpower reporting system was amended in 1976 and by 1977, the system was reporting data with over 95 percent accuracy.

C. THE MARINE CORPS OFFICER ATTRITION AND PROMOTION STRUCTURE

An analysis of a Marine Corps officer as he moves through his career reveals key periods in which an officer

is most likely to leave the service. These periods coincide with promotion points in the service. Under normal circumstances, unrestricted 2nd Lieutenants are promoted to 1st Lieutenants after 24 months in grade. 1st Lieutenants are promoted to Captains in the Fifth year of commissioned service, Captains to Majors during the tenth year, Majors to Lieutenant Colonels during the sixteenth year of service, and Lieutenant Colonels to Colonels during the twenty second year of service. An important influencing factor is the fact that retirement benefits are not achieved until the twentieth year of total service.

1. Lieutenants

Lieutenants normally have a four or five year initial term of service depending on their source of commission and Military Occupational Specialty (MOS), either air or ground. Aviation Lieutenants are some what special. Only a small number of aviation Lieutenants leave the service. This is due to their longer initial term of service, which causes them to reach the grade of Captain before the level of natural attrition is more likely to occur. Attrition of Aviation Lieutenants is primarily a function of accidents, illness, and disciplinary problems.

2. Captains

There is a much greater period of flexibility for Captains given that they are regular officers or reserves with an extension of service agreement. Captains are promoted at five years of commissioned service and are either passed over or promoted to Major by their eleventh year of service. Thus in a period of over six years, an individual may leave the service voluntarily at any point.

D. ATTRITION OF AVIATORS

The attrition rate of aviation officers was affected by the intiation of Aviation Officer Continuation Pay (AOCP) in 1981. AOCP provides a bonus per year to aviation officers and this in turn obligates continued service. The program was applied to all ranks provided the individual met certain active duty flight status requirements. This action by DOD has had its desired effect upon the retention of aviation officers. This accounts, in part, for the reduction in attrition rates for Lieutenant Colonels in 1983 and for Majors and Captains in 1982 and 1983.

E. TERMS

The following terms will be utilized within this project:

1. Aggregation - collection of historical officer data over distinguishing characteristics. For example, the aggregation of all grades within a Military Occupational Specialty (MOS).
2. Attrition - any departure from the Marine Corps by an officer.
3. Attrition rate - idealized ratio of attritions to inventory.
4. End-strength - total number of the Marine Corps officers at the end of a specified period of time.
5. Limited Duty Officer (LDO) - professionally qualified officers specifically designated for limited duty within certain MOS's.
6. Unrestricted - all officers not designated as Limited Duty Officers (LDO).
7. Regular - unrestricted officers not specifically assigned to a "Reserve" status.

8. Reserve - unrestricted officers with a specifically assigned "Reserve" status.

II. STRUCTURE OF THE DATA BASE

A. SUMMARY DATA FILE

The data used in this project was obtained from a summary data file from the Commander, Navy Personnel Research and Development Center (NPRDC). The summary data file was generated from two source files: the Headquarters Master File (HMF) and the Quarterly Statistical Transaction File (STATS). The Headquarters Master File (HMF) was used to produce historical officer inventories as of the beginning of the fiscal year. Inventories were generated for each fiscal year from 1977 through fiscal year 1983. The inventories were identified by distinct characteristics. These characteristics were Military Occupational Specialty (MOS), grade and length of service.

The data in hand does not distinguish between unrestricted officers and limited duty officers. For that reason, discussion of the characteristics that separate these two types cannot be used advantageously herein. Such discussion is included however in anticipation of receiving a more discriminating data base in the future.

The Quarterly Statistical Transaction File was used to generate historical losses. Within this file, if the Type Transaction Code indicated a loss, then the Effective Date of Action field would specify the year and month of the loss. Losses were classified into eight categories for each fiscal year 1977 through 1983. The losses were further classified into distinct elements by MOS, grade and length of service.

B. SUMMARY DATA FORMAT

The summary data file classified the Marine Corps Officer inventory into 40 Military Occupation Specialties, 10 grade levels, 31 lengths of service, and 8 loss categories. The data format is defined by Tables I and II. The column containing the letters A through E refer to the structural zero problem discussed in the next section. See the material surrounding the contents of Table V.

Throughout the remainder of this project, when reference is made to a particular Military Occupational Specialty (MOS), the data code reference from Table II will be used instead of the actual MOS. For example, this project will refer to the Utilities MOS as number 06 and not 11. It should also be understood that the two digit MOS identifier listed in Table II is strictly the Occupational Field identifier in the USMC MOS Manual. This results in MOS, for the purposes of this paper, being an aggregation of what is usually understood to be an MOS cell.

C. STRUCTURAL AND SAMPLING ZEROES

There are two kinds of zero inventory cells that must be dealt with:

1. Structural zeroes: A cell whose inventory is always zero because certain grades and length of service combinations should never appear in the Military Occupational Specialty (MOS). For example, a Lieutenant Colonel with 4 years of service in any MOS or an infantry warrant officer in MOS 03 does not exist.
2. Sampling zeroes: according to the sample data, there is no officer in the particular grade and length of service combination for a given fiscal year and Military Occupational Specialty. This is not a permanent condition, but one that occurs by chance.

TABLE I

DATA FORMAT

COLUMN	DATA
1-2	YEAR - (FISCAL) (1977...1983)
3-4	MOS GROUP (0...39)
5-6	GRADE
7-8	LENGTH OF SERVICE (0...30)
9-13	BEGINNING OF YEAR INVENTORY
14-53	LOSS COUNTS (TYPES 1...8)

CODE	GRADE
0	WARRANT OFFICER (W-1)
1	CHIEF WARRANT OFFICER (CWO-2)
2	CHIEF WARRANT OFFICER (CWO-3)
3	CHIEF WARRANT OFFICER (CWO-4)
4	SECOND LIEUTENANT
5	FIRST LIEUTENANT
6	CAPTAIN
7	MAJOR
8	LIEUTENANT COLONEL
9	COLONEL

CODE	TYPE LOSS
1	VOLUNTARY RESIGNATION
2	VOLUNTARY RETIREMENT
3	INVOLUNTARY - DEATH
4	INVOLUNTARY - DISCHARGE
5	INVOLUNTARY - DISABILITY
6	RELEASE FROM ACTIVE DUTY
7	DISABILITY RETIREMENT
8	INVOLUNTARY RETIREMENT

For the purpose of clarity in this paper, the nonstructural zero region will be referred to as being within the Feasible Region. This region is defined by the lower and upper limits of the length of service for each grade. Table III identifies the minimum and maximum years of commissioned service for the Marine Corps officer grades for the

TABLE II

MILITARY OCCUPATIONAL SPECIALTIES (MOS)

DATA CODE	MOS	CAT	MOS TITLE
00	UN	A	UNKNOWN
01	01	A	PERSONNEL AND ADMINISTRATION
02	02	A	INTELLIGENCE
03	03	C	INFANTRY
04	04	A	LOGISTICS
05	08	A	FIELD ARTILLERY
06	11	D	UTILITIES
07	13	A	ENGINEER, CONSTRUCTION, AND EQUIPMENT
08	14	D	DRAFTING, SURVEYING, AND MAPPING
09	15	D	PRINTING AND REPRODUCTION
10	18	C	TANK AND AMPHIBIAN TRACTOR
11	21	A	ORDNANCE
12	23	B	AMMUNITION AND EXPLOSIVE ORDNANCE DISPOSAL
13	25	A	OPERATIONAL COMMUNICATIONS
14	26	A	SIGNALS INTELLIGENCE/GROUND ELECTRONIC WARFARE
15	28	B	DATA/COMMUNICATIONS MAINTENANCE
16	30	A	SUPPLY ADMINISTRATION AND OPERATIONS
17	31	A	TRANSPORTATION
18	33	A	FOOD SERVICE
19	34	A	AUDITING, FINANCE, AND ACCOUNTING
20	35	A	MOTOR TRANSPORT
21	40	A	DATA SYSTEMS
22	41	B	MARINE CORPS EXCHANGE
23	43	A	PUBLIC AFFAIRS
24	44	A	LEGAL SERVICES
25	46	A	TRAINING AND AUDIOVISUAL SUPPORT
26	55	E	BAND
27	57	D	NUCLEAR, BIOLOGICAL, AND CHEMICAL
28	58	A	MILITARY POLICE AND CORRECTIONS
29	59	B	ELECTRONICS MAINTENANCE
30	60	A	60XX
31	61	A	AIRCRAFT MAINTENANCE
32	63	B	AVIONICS
33	65	B	AVIATION ORDNANCE
34	68	B	WEATHER SERVICE
35	70	D	AIRFIELD SERVICES
36	72	A	AIR CONTROL, AIR SUPPORT, AND ANTI-AIR WARFARE
37	73	A	AIR TRAFFIC CONTROL
38	75	C	PILOTS AND NAVAL FLIGHT OFFICERS
39	99	E	IDENTIFYING MOS AND REPORTING MOS

unrestricted as well as the Limited Duty Officer. Recall, the latter type is not separated in the current data.

TABLE III

MINIMUM AND MAXIMUM YEARS-OF-COMMISSIONED SERVICE FOR OFFICER GRADES

Grade	Unrestricted MIN YCS	Unrestricted MAX YCS	LDO MIN YCS	LDO MAX YCS
WO1	1	4	-	-
CWO2	4	9	-	-
CWO3	8	13	-	-
CWO4	12	24	-	-
2LT	1	4	-	-
1LT	1	8	1	8
CAPT	4	13	2	12
MAJ	9	20	5	18
LTCOL	14	23	8	23
COL	21	30	-	-

In order to get the officer's length of service one must add the officer's prior enlistment period to the officer's years of commissioned service. Since prospective warrant officers are required to have a minimum of five years of service prior to consideration by the officer selection board, the logical lower limit for length of service for warrant officers is five years. The LOS then would include the total of this service and the respective years of commissioned service. Additionally, warrant officers are required to have 10 years of service prior to consideration to become Limited Duty Officers. This affects the lower limit of the length of service for Limited Duty Officers. The logical lower limit for length of service for LDO's would be 10 years.

The summary data provided for this project did not distinguish between unrestricted and Limited Duty Officers. Additionally, the length of service was the characteristic provided instead of the years of commissioned service for each of the officer grades. Table IV identifies the lower

and upper limits of the Feasible Region of the length of service for Marine Corps officer grades utilized in this project.

TABLE IV

MINIMUM AND MAXIMUM LENGTH OF SERVICE FOR OFFICER GRADES

Grade	MIN LOS	MAX LOS
WO1	5	20
CWO2	8	30
CWO3	12	30
CWO4	16	30
2LT	1	20
1LT	1	20
CAPT	4	30
MAJ	9	30
LTCOL	14	30
COL	21	30

Note the maximum limit in the majority of grades can feasibly be 30 years. This was substantiated by the actual data wherein the maximum length of service for grades CWO3 through Colonel was 30 years. The maximum length of service for WO1's, 2LT's and 1LT's was limited to 20 years because of the very small number of actual officers with 21 to 30 years of service within these grades.

One method of considering the size of the feasible region is to create a multidimensional array using all combinations of different characteristics within the data. If all characteristics were considered for all seven fiscal years, the size of the multidimensional matrix would be 7x40x31x10x8. This would represent 7 years, 40 MOS's, 31 LOS's, 10 grades and 8 loss types. This would create 694,400 different cells that could be dedicated to loss

counts. However, to bring the dimension to a more manageable level, consider one year of data for each loss type having a cell count of 40x31x10. Therefore, for each loss type within a fiscal year there is a total of 12,400 cells. Many of these cells are infeasible and are designated those with structural zeroes.

The region of structural zeroes is the area outside of the minimum and maximum length of service for the officer grades as listed in Table IV. But it is further articulated because of additional constraints imposed by the various MOS classifications. This is largely because of the suitability of certain MOS's being assigned to unrestricted, warrant and Limited Duty Officers. The particular MOS's are identified for illustration purposes by letter in Table II according to the general categorization listed in Table V. This general MOS categorization specifies by category the inclusive grades within certain MOS's, the number of MOS's with the specified grade structure, the number of structural zeroes per MOS, and the total structural zeroes within each category. The total number of structural zeroes is 6149. Therefore, the number of cells within the feasible region is 6251 for each loss type for a given fiscal year.

D. CENTRAL DATA

The following is an observation made by the author regarding the summary data utilized in this project. This data, which was provided by Navy Personnel Research and Development Center (NPRDC), was classified as central data according to Bartholomew [Ref. 1: p. 25]. Bartholomew defines central wastage (attrition) rates as follows:

1. The central wastage (attrition) rate is the number of leavers during the period who were in this class when they left divided by the average number in this class during the period.

TABLE V

STRUCTURAL ZEROES CATEGORIES

Category	Grades Within MOS	Number of MOS	Stru. Zeroes per MOS	Total Zeroes per Cat.
A	WO1...LTCOL	23	129	2967
B	WO1...CWO4,LDO	8	159	1272
C	2LT...LTCOL	3	202	606
D	WO1...CWO4	5	237	1185
E	WO1...COL	1	119	119
TOTAL		40		6149

A problem arose on several occasions when the data was disaggregated to a level where the inventory was very small. For example, when examining the inventory in a particular fiscal year, the inventory was zero for a Length of Service (LOS) and Military Occupational Specialty (MOS) combination. Examining the inventory in the next fiscal year for the same LOS and MOS combination also was zero. However, the problem arises when the number of leavers is equal to or greater than one. The average of the inventories for the two fiscal years is zero. Using this result in the estimation of the attrition rate would be ambiguous. The zero in the denominator of the ratio of leavers to average inventory results in an undefined expression.

For the purpose of removing this ambiguity in the data, the following policy was adopted to define the central inventory number for the officer force at disaggregated levels: For any cell or collection of cells

1. Let, t = 1...6, refer to the years 1977...1982.
2. Let, Y(t) = Number of losses in year t.
3. Let, INV(t) = Inventory in the beginning of year t.

4. Let, $N(t)$ = Maximum of $Y(t)$ and the average inventory in year t, using the beginning inventory in year t and t+1 and computing their average: $(INV(t)+INV(t+1))/2$.
5. Let, m = Sum of $Y(t)$ divided by sum of $N(t)$, (both sums over t) represents the empirical central attrition rates for the particular cell or aggregate.

The data bank that will ultimately be utilized in the parent project at NPRDC, of which this project is a subset, will contain the appropriate inventory and attrition data so that the above modifications will not be required. It is currently planned, by NPRDC, to produce the historical officer inventory as of the beginning of a month and the losses will be identified by the month in which the officer leaves the Marine Corps. This will provide the elements for a more accurate estimation of the attrition rate on the disaggregate level.

III. CURRENT SYSTEM

A. OVERVIEW

In order to gain an appreciation for the complexity of the Manpower Planning process, the author during his experience tour at Headquarters Marine Corps had the opportunity to contact key individuals responsible for the execution of the models currently utilized within the Manpower department. The methodology used to identify the current system included interviews with the action officers of the Officers Plans Section (Code MPP-30), briefings and interviews with representatives from Allocations Section of Manpower Control Branch (Code MPC-20), Officer Assignment Branch (Code MMOA), Manpower Plans, Programs and Budget Section (Code MPP-40), Manpower Management Information Systems Branch (Code MPI), Assistant Deputy Chief of Staff for Manpower (Code M), and examination of existing manuals, Requirements Statement for Officer Planning System, and Statement of Work for the Development of the Officer Inventory and Manpower Flow Data Base System.

The goal of the officer planning process is to shape the officer force structure over the 7 planning years (current, budget, and five Program Objective Memorandum (POM) years). This force structure is affected by three influencing factors: existing inventory, projected losses, and expected gains. The Officer Plans Section (Code MPP-30) supports the Manpower Department at Headquarters Marine Corps by preparing plans that address each of these factors.

The plans which input the target inventory calculations are the Promotion Plan, the Lateral Move and Directed Lateral Move Plans, the Inventory Projection, and the Manpower Plan.

Losses in the officer force are the result of retirements, resignations, releases, discharges and other factors (e.g., death). Among these factors, releases and mandatory (statutory) retirements addressed in the Oficer Inventory Control Plan, can be estimated with a high degree of certainty. The remaining loss factors are estimations based on forecasted attrition rates utilizing historical data. All of the factors contributing to losses in the officer inventory are addressed in the Inventory Projection Loss Plan.

Gains in officer force are the result of accessions and augmentation. Accessions are identified in the Accession Plan and augmentations are identified in the Augmentation Plan. The number of field grade officers are directly influenced by statutory limitations specified in Title 10, U.S. Code. These limitations affect both of these plans. In the planning process, in order to plan for an adequate number of field grade officers in the future, one must currently access a sufficient number of second lieutenants.

The force structure is also affected by the mixture of skills of the officer inventory. The plans used to support the required mix and levels of skills are the TBS/MOS Distribution Plan, the Entry Level Training Plan, the Lateral Mover Plan, the Continuation Plan, and the Promotion Plan.

B. SOURCES OF INFORMATION

Input from the following external sources provide information required by the Officer Plans Section to create the plans identified above.

1. Statutory limitations specified in Title 10, U.S. Code.

2. Desired inventory force structure as enumerated in the Grade Adjusted Recapitulation (GAR).
3. Feedback data from agencies which execute the plan (for example, actual monthly promotions).
4. Historical data from JUMPS/MMS on officer service characteristics and any changes in the officer inventory over time.
5. Current data on the existing officer inventory utilizing the Officer Slate file and the Lineal List file.

C. FLOW OF INFORMATION

The planning process consists of three major elements comprised of several processes. These elements are:

1. Preparation for planning.
2. Plan development.
3. Force structure analysis.

The preparation process includes development of the targeted force structure to include distribution of officers by skill, sex, unrestricted/restricted categories, and grades based on historical flow rates and USMC policy objectives. The requirements definition process establishes the authorized number of officers in each Military Occupational Specialty (MOS) and grade for the five planning years and determines the target force. The programmed requirement is specified in the Numerically Adjusted Recapitulation (NAR).

The planning process also includes the determination of officer attrition rates. Currently the attrition rate determination method is accomplished in conjunction with the Officer Promotion Planning Process. This model divides the attrition rates into two categories: statutory and non-statutory. The statutory attrition rate results from USMC and DOD policies and is applied directly to the officer

inventory. The non-statutory rate is derived from a time series analysis of officer inventory. This rate is based on historical data and assumes continuation of present trends. Currently there is no attempt to provide predictive rates based on econometric techniques. However, such an approach is being researched by the Navy Personnel Research and Development Center for use by the Manpower Department.

The development process includes projection of the officer inventory and generation of various plans. The goal of the inventory projection process is to provide a breakdown by month of the number of officers in each grade, Military Occupational Specialty (MOS), and other officer attributes (for example, restricted, years of commissioned service, sex, etc.). This breakdown is for each of the planning years. Currently the inventory projection is determined in one of several ways, depending on the plan being developed. The Promotion Planning Process projection is utilized if the level of aggregation is adequate for the plan. Otherwise, manual calculations are used to aggregate from the promotion planning process inventory projection or to generate a new projection.

The plan generation process is complicated by the interdependence of the plans and the timing of their production. Table VI identifies the plan generation process including each plan, the frequency of the generation and update of each plan, the users, and the sources of data.

The force structure analysis process involves verification of the generated plans and evaluation of their quality. Comparison is made of the results of the planning process with the USMC objective force specified by the Grade Adjusted Recapitulation (GAR) and the promotion flow points specified by Defense Officer Personnel Management Act (DOPMA).

TABLE VI

OFFICER PLAN GENERATION SUMMARY

PLAN	GENERATED	UPDATED	USED BY	SOURCES OF DATA
Promotion plan	Annual	Monthly	Prom. Brd MPP-30 RES MMPR	1,2,3,4, 10
Augmentation Plan	Annual	Monthly	MPP-30 MMOA ORB	1,2,3,4, 11
Inventory Projection	Annual	Annual	MPP-30	1,2,3,4, 5,6,7,8, 9
Inventory Proj. Loss Plan	Annual	Annual	MPP-30	1,2,7
Continuation Plan	Annual	Annual	MPP-30	1,2,4,5
Accession Plan	Annual	Monthly	MPP-30 MMRD	1,2,3,4, 6,8
Entry Level Training Plan	Annual	Monthly	MMOA TBS T	1,2,4
TBS/MOS Distribution Plan	Annual	Monthly	MMOA TBS T	1,2,4
Lateral Move Plan	Annual	Monthly	MPP-30 MMOA	1,2,4,5
Manpower Plan	Annual	Monthly	MPP-30 MPP-40	1,2,3,4, 5,6,8,9
Officer Control Plan	Annual	Monthly	MPP-30	1,2,4,5

Source of Data Codes

1. HMF
2. Attrition Rates
3. DOPMA
4. GAR
5. Budget
6. Candidate List
7. JUMPS/MMS
8. WO Accessions
9. Retirement/Separation
10. Lineal List
11. Officer Slate File

D. PLANS

Promotion plan is generated by the use of the Promotion Planning Process model. The model's output is a statistical projection of the officer inventory over a period of up to 10 years. It also provides several built in promotion strategies.

Augmentation Plan is generated to provide the Officer Retention Board (ORB) with information on the vacancies available in each fiscal year of the planning process regarding officers in a category (i.e., Aviation, Ground, and Naval Flight Officer) with a certain number of years of commissioned service.

Inventory Projection Plan is generated for a five year projection and is specified for each year by grade, years of commissioned service, and category (unrestricted or Limited Duty Officer (LDO)).

Inventory Projection Loss Plan contains data on expected losses from the inventory projection process. The plan is utilized to indicate which year of commissioned service the losses are most likely to occur.

Officer Inventory Control Plan provides information on the number of reserve officers whose contracts expires during the fiscal year. It is influenced by the Augmentation Plan and provides guidance on the number of the officers who may be retained.

Continuation Plan specifies the number of officers that may be granted an extension of active duty depending upon their Military Occupational Specialty (MOS).

Accession Plan identifies the desired number of incoming potential officers and receives as input the number of vacancies for each year of the planning process. Since accessions into the USMC are by officer candidates who have usually signed contracts several years prior to graduation,

the pool of accessions for the first three years are generally well known. The goals for the remaining planning years are established on projected officer losses and on budgeted end strengths.

Entry Level Training Plan and the TBS/MOS Distribution Plan ensure that the entry level officers have the mix of Military Occupational Specialties that is needed to fully support the force structure.

Lateral Move Plan provides information on the shortage or overage in each Military Occupational Specialty by year-group.

Manpower Plan is a summary of gains and losses in each grade.

IV. MARINE CORPS OFFICER ATTRITION MODEL

A. DEFINITION

A definition for a manpower model offered by Bartholomew is:

> A manpower model is a mathematical description of how change takes place in the system. First of all this requires the specification of any constraints under which the system operates . . . . Secondly, a model must specify the mechanism which generates flows. Some flows, such as promotion or demotion, are under the direct control of management . . . . Other flows, such as voluntary wastage, are not under direct management control and assumptions about their future levels is likely to be based on a blend of historical data and management judgement. [Ref. 1: p. 7]

B. ASSUMPTIONS

Assumptions regarding flows of personnel within a system can be classified in various ways. If a manpower policy dictates that a specific percentage of individuals in a particular category would leave in a given time period then the assumption about the flow would be deterministic. Given a level of inventory in the category under the deterministic assumption, then there is no uncertainty about how many individuals will actually leave. A deterministic model contains no random variables and there is a unique set of model output data for a given set of inputs. [Ref. 3: p. 3]

However, if each individual in a category had a certain probability of leaving in a given time period, then the assumption about the flow would be stochastic. Given a level of inventory in the category under the stochastic assumption, then a prediction of the number of individuals

leaving would not be precise. A stochastic model contains one or more random variables. The output of a stochastic model is random and thus only estimates the true characteristics of the system [Ref. 3: p. 3]. The output of a stochastic model is a probability distribution. If individuals within a system behave independently regarding their decision of leaving, the actual number of leavers has a binomial distribution whose average is a certain percentage of the inventory within a category [Ref. 1: p. 7]. Voluntary attrition from the Marine Corps is a stochastic flow since it results from a number of more or less independent individual decisions.

C. ATTRITION RATES

A fundamental role in manpower analysis is played by attrition rates. The time period can be any duration. The most commonly used periods are months and years. The use of the central attrition rate when the categories are defined with respect to length of service can be considered as estimating the constant rate of leaving or separation during a year. Using this method, the standard error of the central attrition rate can be estimated if some assumptions are made about the attrition process:

1. The number in the category throughout the year was constant and equal to the average number in this class during the year.
2. Let, m = the central attrition rate.
3. Let, S* = Average number in a category during the year.
4. Let, L* = Number of leavers during the year who were in this category when they left.
5. Then, m = L* / S*.

6. Each individual in the category is subject to the probability (mdx) of leaving in each interval (x,x+dx). [Ref. 1: p. 25]

Under these assumptions the number of losses would be distributed Poisson (S*m). The estimated standard deviation of the central attrition rate would be: [Ref. 1: p. 25]

$$sd(m) = m / sqrt(L^*)$$

D. MARINE CORPS OFFICER CENTRAL ATTRITION RATES

Some empirical attrition rates are provided for the purpose of familiarization. First are macro (large aggregate) examples and proceed progressively to more refined categorized examples. An example of the estimated central attrition rates and their estimated standard deviations are provided in Table VII. The average inventory over the seven year period is also provided in Table VII in order to illustrate the number of officers within each MOS.

The rates in Table VII are an aggregation of all grades within a Military Occupational Specialty (MOS) and all loss types. The time period includes all seven years of inventory counts and the losses for fiscal years 1977 to 1982. The central attrition rate was calculated as follows:

1. Let, t = 1...6, covering years 1977...1982.
2. Let, Y(t) = Number of losses in year t.
3. Let, INV(t) = Inventory in the beginning of year t.
4. Let, N(t) = Average inventory in year t, using the beginning inventory in year t and t+1 and computing their average: (INV(t)+INV(t+1))/2.
5. Let, m = Sum of all Y(t) divided by sum of all N(t).

The central attrition for each Military Occupational Specialty is illustrated for the length of service period from 0 to 30 in Appendix A, in Figures A.1 toA.10. These figures illustrate the aggregate central attrition rate for

TABLE VII

CENTRAL ATTRITION RATES AND STANDARD DEVIATIONS

MOS	AVERAGE INVENTORY	RATES	STANDARD DEVIATIONS
01	531	.121	.006
02	288	.119	.008
03	3059	.091	.002
04	183	.104	.010
05	1322	.101	.004
06	36	.080	.019
07	657	.114	.005
08	5	.034	.034
09	6	.027	.027
10	512	.105	.006
11	102	.090	.012
12	76	.100	.015
13	782	.138	.005
14	128	.099	.011
15	94	.098	.013
16	1148	.139	.004
17	26	.156	.032
18	40	.120	.022
19	309	.109	.008
20	364	.115	.008
21	251	.143	.010
22	22	.156	.035
23	59	.150	.020
24	459	.193	.008
25	24	.124	.029
26	17	.081	.029
27	11	.047	.033
28	156	.093	.010
29	119	.120	.013
30	262	.100	.008
31	0	.000	.000
32	65	.042	.017
33	89	.067	.011
34	14	.108	.036
35	32	.088	.021
36	528	.125	.006
37	108	.091	.012
38	5644	.081	.002
39	1296	.100	.004
TOTAL	18780	.102	.001

all eight types of losses over the six years from 1977 to 1982. These figures demonstrate the breakpoints and discontinuities within each MOS over the distribution of length of service. One can very easily note the increased rates which

occur at certain years of length of service. The increased rates frequently occur at the four, 20, and 30 years of length of service. The central attrition rate for each MOS is illustrated for the grades from Warrant Officer to Colonel (labelled 0...9 on the abscissa) in Appendix A, in Figures A.11 to A.20. These figures also illustrate the aggregate central attrition rate for all eight types of losses over the same six years.

The central attrition rate for each fiscal year aggregated over all categories is provided in Table VIII and illustrated in Figure 4.1. The trend is stable for the first three years and a decreasing trend is evidenced in the last three years.

TABLE VIII

USMC OFFICER LOSS RATES PER FISCAL YEAR

1977	1978	1978	1980	1981	1982
.1133	.1137	.1136	.1012	.0928	.0781

Table IX is provided in order to gain an appreciation of the number of losses within each loss type. This table includes the number of losses for each fiscal year 1977 through 1983 aggregated over all grades, LOS, and MOS. This table also identifies the loss types. One can note the large number of losses throughout the seven years in types two, four, and six. Additionally, after reaching a total high in 1978, the general trend is a decreasing number of losses.

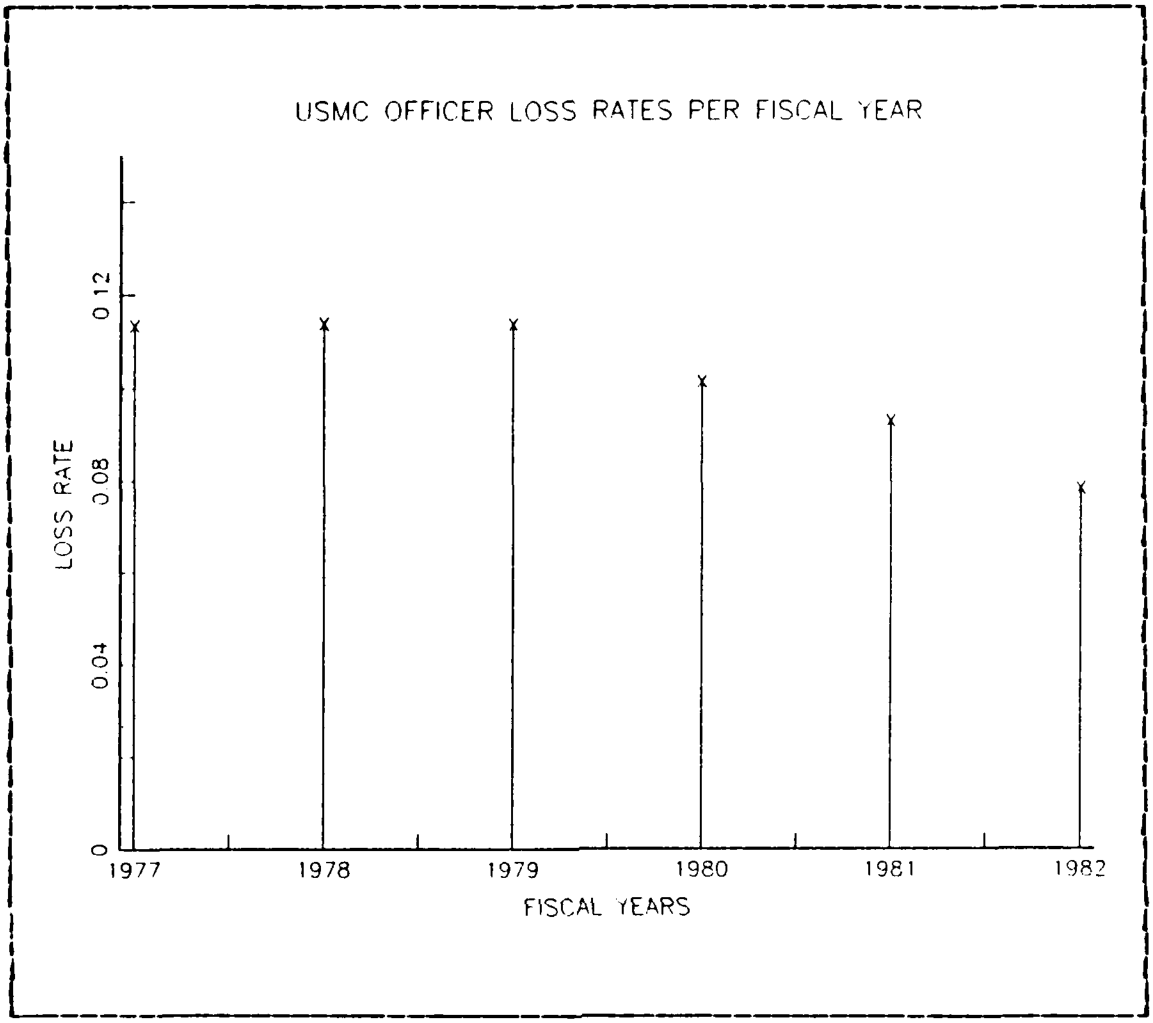


Figure 4.1 USMC Officer Loss Rates Per Fiscal Year.

The central attrition rates for this level of aggregation for fiscal years 1977 through 1982 are illustrated in Table X.

The central attrition rates for each Military Occupational Specialty (MOS) and for each fiscal year 1977 through 1982 is provided in Table XI. These rates are aggregated over all loss types. These rates are graphically illustrated in Appendix A, in Figures A.21 to A.28. One can observe the change of rates over time for each MOS. It can

TABLE IX

LOSSES PER FISCAL YEAR BY LOSS TYPE

LOSS TYPE	1977	1978	1979	1980	1981	1982	1983	TOTAL
L1	118	90	109	40	24	37	35	453
L2	675	701	681	594	531	524	362	4068
L3	32	42	38	39	34	22	27	234
L4	364	515	536	572	532	431	317	3267
L5	1	1	0	0	1	0	0	3
L6	894	772	750	595	522	406	436	4375
L7	15	27	17	2	3	2	10	76
L8	52	50	39	35	43	30	41	290
TOTAL	2151	2198	2170	1877	1690	1452	1228	12766

LOSS TYPES

L1: Voluntary Resignation
L2: Voluntary Retirement
L3: Involuntary - Death
L4: Involuntary Discharge
L5: Involuntary Disability
L6: Release from Active Duty
L7: Disability Retirement
L8: Involuntary Retirement

be noted with interest that the MOS's with larger inventories (refer to Table VII for the average inventory within each MOS) do not change as much over time as the MOS's with smaller inventories. For example, the infantry MOS 03 has the second largest inventory count. The range of the rates over time fluctuate from a low of 0.070 to a high of 0.106 giving a range of 0.036. The logistic MOS 04, on the other hand, has a much smaller inventory count and the rates over time fluctuate from a low of 0.049 to a high of 0.22, giving a range of 0.171. In other words, the total inventory count in a category has a definite effect on the change of rates over time.

TABLE X

CENTRAL ATTRITION RATE PER FISCAL YEAR BY LOSS TYPE

LOSS TYPE	1977	1978	1979	1980	1981	1982
L1	.006	.005	.006	.002	.001	.002
L2	.036	.036	.035	.031	.028	.028
L3	.002	.002	.002	.002	.002	.001
L4	.019	.027	.028	.030	.028	.023
L5	.000	.000	.000	.000	.000	.000
L6	.047	.040	.039	.031	.027	.022
L7	.001	.001	.001	.000	.000	.000
L8	.003	.003	.002	.002	.002	.002

The change in the rates are also affected by the level of disaggregation. The central attrition rates for the major Military Occupational Specialty groups (Aviation, Combat Support, Ground Combat and total) are listed in Table XII. These major groups are illustrated in Appendix A, in Figures A.29, A.30, A.31, and A.32. The rates demonstrate the aggregation over all types of losses. For brevity the illustrations were limited to the grades of Second Lieutenant to Lieutenant Colonel.

A further disaggregation is listed in Table XIII and illustrated in Figure 4.2. These rates demonstrate the aggregation within the Combat Support group over all lengths of service (LOS) for the voluntary retirements (type 2 loss) for the grades of First Lieutenant, Captain, Major and Lieutenant Colonel. A common observation within this loss type is the increased rate as the grade increases. This phenomenon also corresponds to the increased rate of attrition in the larger LOS categories.

Another level of disaggregation of the Combat Support group is listed in Table XIV and illustrated in Figure 4.3. These rates demonstrate the aggregation within the Motor

TABLE XI

CENTRAL ATTRITION RATES FOR FISCAL YEARS BY MOS

MOS	1977	1978	1979	1980	1981	1982
1	.132	.137	.134	.132	.105	.092
2	.136	.151	.111	.112	.120	.093
3	.106	.092	.098	.088	.086	.070
4	.049	.220	.079	.097	.128	.085
5	.110	.103	.120	.118	.079	.073
6	.057	.121	.131	.029	.133	.024
7	.128	.119	.113	.109	.124	.087
8	.286	.000	.000	.000	.000	.000
9	.000	.000	.200	.000	.000	.000
10	.135	.097	.150	.094	.091	.058
11	.109	.078	.110	.103	.078	.065
12	.068	.126	.113	.132	.079	.079
13	.131	.152	.153	.145	.132	.112
14	.128	.158	.134	.079	.068	.076
15	.127	.167	.125	.115	.049	.043
16	.162	.139	.160	.142	.111	.113
17	.172	.151	.208	.120	.077	.213
18	.123	.173	.100	.099	.146	.076
19	.164	.113	.096	.112	.084	.107
20	.105	.137	.108	.148	.116	.081
21	.123	.178	.140	.170	.130	.118
22	.192	.078	.250	.270	.063	.056
23	.211	.192	.087	.164	.113	.171
24	.244	.209	.195	.192	.154	.148
25	.133	.130	.078	.145	.122	.133
26	.065	.067	.194	.059	.056	.056
27	.000	.000	.000	.000	.000	.091
28	.093	.102	.062	.115	.104	.083
29	.106	.165	.137	.140	.113	.049
30	.086	.119	.083	.101	.120	.089
31	.000	.000	.000	.000	.000	.000
32	.000	.000	.000	.000	.065	.031
33	.072	.047	.044	.086	.043	.110
34	.148	.000	.207	.074	.074	.148
35	.151	.066	.030	.114	.029	.154
36	.159	.139	.139	.098	.111	.094
37	.165	.113	.045	.103	.051	.076
38	.086	.102	.093	.075	.073	.056
39	.090	.093	.123	.089	.114	.094

Transportation MOS over all LOS for the voluntary retirements (type 2 loss) for the grades of First Lieutenant, Captain, Major, and Lieutenant Colonel. This illustrates the irregular distribution resulting from voluntary attrition. The number of losses, even with aggregation over LOS, is relatively small. The actual losses in this table range from zero to seven.

TABLE XII

CENTRAL ATTRITION RATES

AVIATION (MOS 38)

GRADE	1977	1978	1979	1980	1981	1982
2LT	.030	.020	.031	.026	.006	.017
1LT	.055	.082	.032	.025	.029	.018
CAPT	.148	.173	.201	.157	.163	.096
MAJ	.068	.077	.069	.049	.047	.045
LTCOL	.125	.171	.144	.145	.108	.145

COMBAT SUPPORT GROUP (MOS 07,13,20)

GRADE	1977	1978	1979	1980	1981	1982
2LT	.003	.025	.023	.009	.010	.015
1LT	.249	.240	.232	.187	.197	.142
CAPT	.082	.082	.104	.125	.106	.080
MAJ	.123	.119	.107	.145	.125	.080
LTCOL	.136	.193	.114	.155	.154	.114

GROUND COMBAT GROUP (MOS 03,05,10)

GRADE	1977	1978	1979	1980	1981	1982
2LT	.010	.016	.011	.008	.013	.011
1LT	.231	.190	.220	.151	.137	.121
CAPT	.081	.059	.074	.093	.090	.065
MAJ	.074	.051	.060	.060	.059	.032
LTCOL	.112	.129	.131	.139	.100	.109

TOTAL MOS

GRADE	1977	1978	1979	1980	1981	1982
2LT	.035	.032	.038	.025	.020	.026
1LT	.172	.161	.146	.118	.111	.093
CAPT	.129	.128	.140	.133	.123	.088
MAJ	.099	.095	.099	.084	.075	.056
LTCOL	.119	.154	.155	.146	.110	.130

A further disaggregation to include each LOS (not illustrated), would demonstrate the effect of a smaller number of

TABLE XIII

COMBAT SUPPORT GROUP VOL. RET. (LOSS TYPE 2)
CENTRAL ATTRITION RATES

GRADE	1977	1978	1979	1980	1981	1982
1LT	.004	.005	.000	.004	.002	.004
CAPT	.031	.017	.023	.017	.021	.010
MAJ	.103	.105	.092	.113	.090	.068
LTCOL	.122	.178	.114	.147	.145	.114

elements within a cell. This is particularly evident when disaggregation is within a sparsely populated MOS and the desired rate is at a particular grade, LOS, and loss type combination.

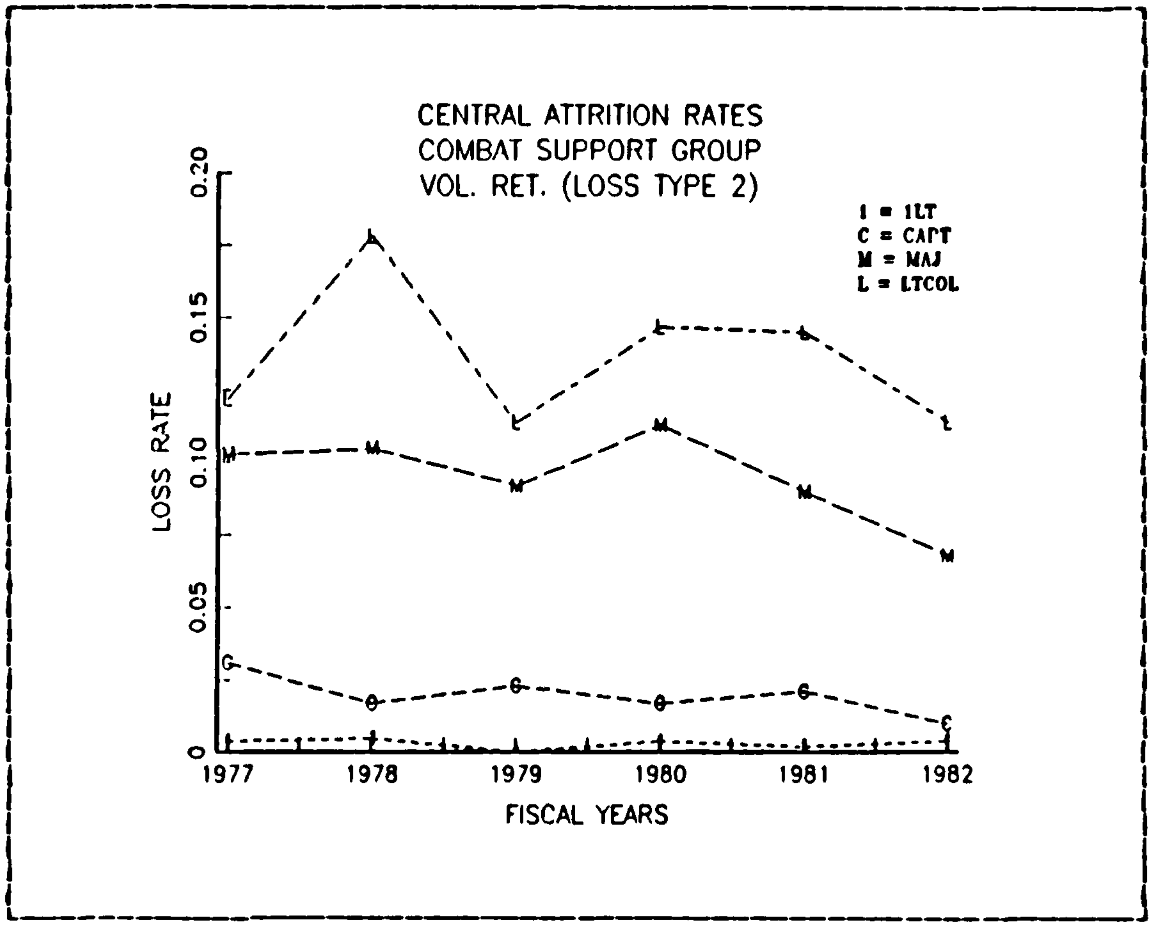


Figure 4.2 Combat Support Group Vol. Ret. (Loss Type 2) Central Attrition Rates.

TABLE XIV

MOTOR TRANSPORT (MOS 20) VOL. RET. (LOSS TYPE 2)
CENTRAL ATTRITION RATES

GRADE	1977	1978	1979	1980	1981	1982
1LT	.000	.023	.000	.011	.000	.000
CAPT	.164	.049	.043	.047	.022	.021
MAJ	.000	.298	.169	.235	.041	.107
LTCOL	.000	.857	.286	.222	.000	.000

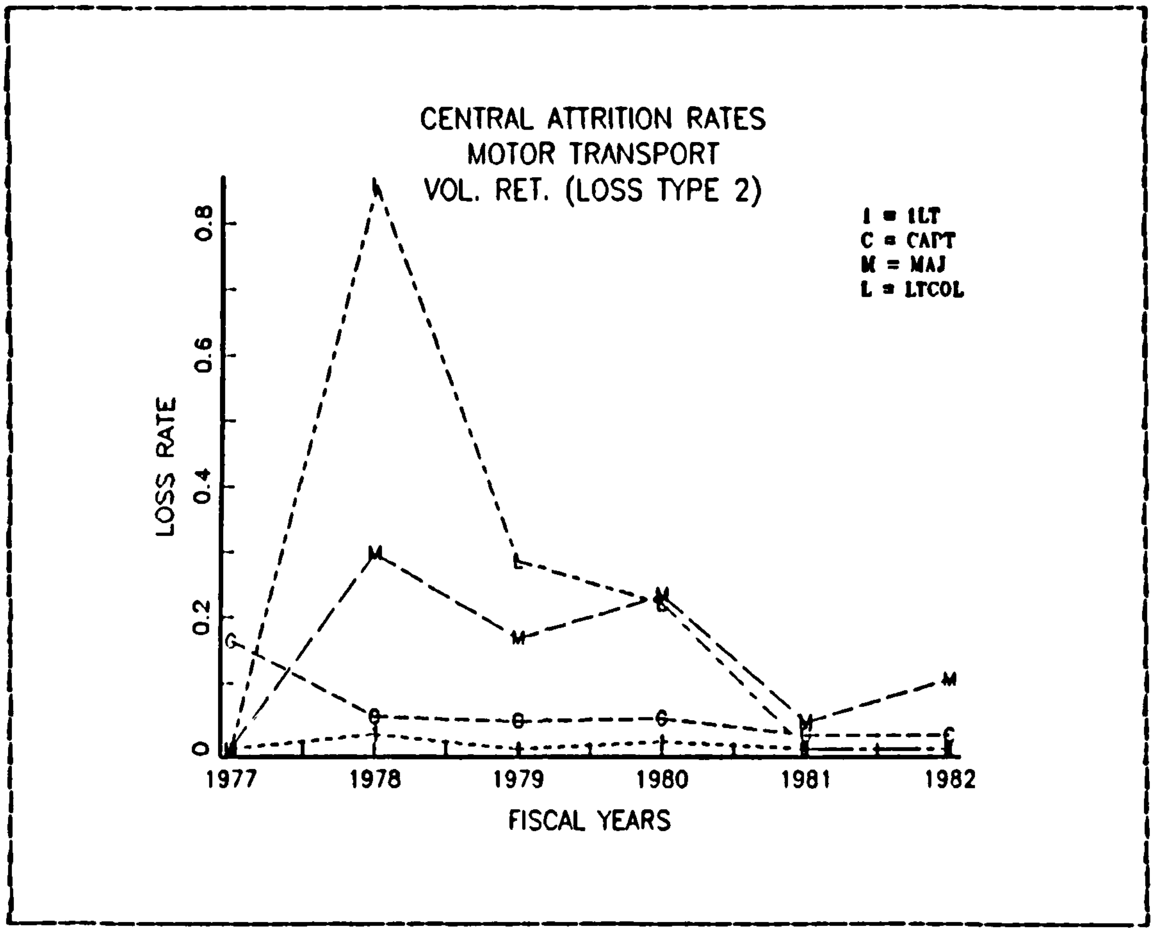


Figure 4.3 Motor Transport Vol. Ret. (Loss Type 2) Central Attrition Rates.

V. JAMES-STEIN TECHNIQUES AND VALIDATION PROCEDURES

A. BACKGROUND

The first basic process in statistics is the simple act of counting and the second most basic process is averaging. One use of observed averages is to predict or estimate unobserved quantities. For example, a baseball player who gets eight hits in 25 times at bat is said to have a .320 batting average, which is the ratio of 8 to 25. In computing this statistic an estimate is formed of the player's true batting ability in terms of his observed average rate of success at the plate. A common reply to the inquiry regarding this baseball player's next 100 times to bat would be that he would probably get 32 more hits. In traditional statistical theory, no other estimation rule is uniformly better than the observed average. The attrition in cells is modeled as a set of Bernoulli trials whose number is the inventory of that cell. The process of choosing leavers is viewed as independent from individual to individual. A convient anology might be that of a baseball players batting average in which each time at bat is viewed as an independent Bernoulli trial.

Using the above Bernoulli (Binomial) model, the rates (or batting averages) are the Maximum Likelihood estimators of the Bernoulli parameter. The maximum likelihood procedure examines the likelihood function of the sample values and takes as the estimates of the unknown parameters those values that maximize the likelihood function [Ref. 4: p. 363]. These estimates are optimal (in particular admissible) if we are talking about one cell (or one ball player). But it is no longer true if we are dealing with three or more cells (or ball players).

Charles Stein [Ref. 5] showed in 1955 that it is possible to make a uniform improvement on the maximum likelihood estimator (MLE) in terms of total squared error risk when estimating several means from independent normal populations. In 1961, James and Stein [Ref. 6] presented to the Berkeley Symposium an estimator which shrinks the maximum likelihood estimator toward the origin, and can improve (i.e. lower risk) on the maximum likelihood estimator quite substantially provided there are three or more cell. To continue the baseball example, if one considers three or more baseball players and desires to predict future batting averages for each of them, then there is a procedure that is better than simply extrapolating from the separate averages. Efron and Morris [Ref. 7] demonstrated an application of the Stein rule and its generalizations applied to predict baseball averages, to estimate toxomosis prevalence rates, and to estimate the exact size of Pearson's chi-square test with results from a computer simulation. The result of these applications was a mean square error of the Stein rule being less than half of that when one uses the individual empirical averages.

The first step in applying Stein's method is to determine the average of the averages. This grand average or grand mean must lie between 0 and 1. The essential process in Stein's method is the "shrinking" of all the individual averages toward this grand mean. When the Stein's method is applied to attrition rates, the results are as follows.

1. If an aggregated rate is higher than the grand mean then it will be reduced.
2. If an aggregated rate is lower than the grand mean then it will be increased.

Figure 5.1 is a simple illustration of this "shrinking" concept utilizing fictitious values for the empirical values. In this simple illustration the effective shrinkage is 25 percent.

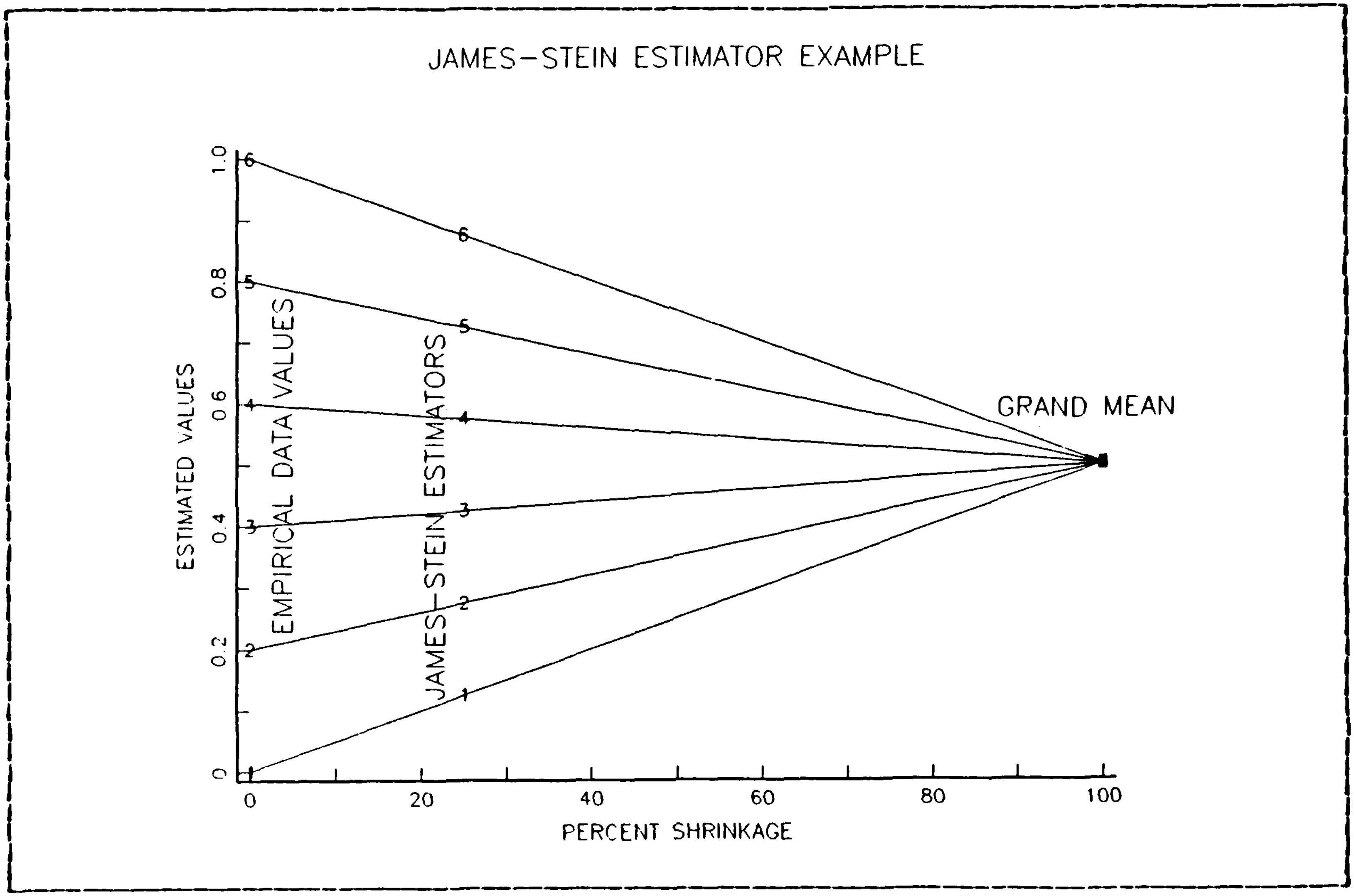


Figure 5.1 James-Stein Shrinkage Example.

B. JAMES-STEIN ESTIMATOR

The James-Stein estimator for each desired cell is found through the following equation.

1. Let Y^* = the grand mean.
2. Let m = sample attrition rate for each cell.
3. Then, $p = Y^* + c(m-Y^*)$, for a specialized value of "c", is the James-Stein estimator.

See Appendix B, for the James-Stein estimation algorithm. See Appendix C, for more mathematical details on the equation used for the James-Stein estimator. The amount $(m-Y^*)$ is the amount by which the sample attrition rate differs from the grand mean of the desired cells. The equation thus states that the James-Stein estimator "p" differs from the grand mean by this same quantity $(m-Y^*)$ multiplied by a coefficient, "c". The coefficient "c" is the "shrinking factor" used in the James-Stein estimation process. If "c" was equal to 1, then the resultant James-Stein estimator for a given cell would be identical to the cell's sample attrition rate. In other words, $p=m$. Stein's theorem [Ref. 6] states that the shrinking factor is always less than 1. Its actual value is determined by the collection of all observed attrition rates.

As an example, if the shrinking factor "c" was equal to .3, then each attrition rate will shrink about 70 percent of the distance to the grand mean. Three examples follow:

1. Assume the sample attrition rate was .10, the grand mean was .05, and "c" equal to .3. Then $p=.05+.3(.10-.05)$. The result is .065. By Stein's theorem, the attrition rate is best estimated by .065 rather than the historical sample attrition rate. Shrinkage was down toward the grand mean.
2. Assume the sample attrition rate was .50, the grand mean was .05, and "c" equal to .3. Then

p=.05+.3(.50-.05). The result is .185. The estimator is smaller than the sample attrition rate. Shrinkage was down toward the grand mean.

3. Assume the sample attrition rate was .05, the grand mean was .10, and "c" equal to .7. Then p=.10+.7(.05-.10). The result is .065. Shrinkage was up toward the grand mean.

There are several expressions for the James-Stein estimator, but all of them have in common a shrinking factor "c". Without exhibiting any of the various formulas for "c", the following is a description of their general behavior. See Table XV for the James-Stein formula utilized in this project. See Appendix C for further background and formulas.

Let "K" be the number of unknown means; let sigma be the within population standard deviation; and let SSB be the sum of square of cell means measured from the grand mean. Holding sigma and "K" constant in the formula for "c", the shrinking factor is affected by SSB. The shrinking factor "c" becomes smaller as SSB gets smaller. The predicted means are more affected by this condition and the resultant mean is closer to the grand mean. On the other hand, as SSB increases, "c" increases and the shrinkage is less drastic. In effect the James-Stein procedure makes a preliminary estimate that all the unobservable means are near the grand mean. If the data supports the condition wherein the sample means are not too far from the grand mean, then the estimates are all shrunk further toward the grand mean. If this condition is contradicted, then the shrinking is minimal. The distribution of the sample means around the grand mean effects the shrinking factor. The number of means being estimated also influences the shrinking factor. If "K" is rather large, the shrinking to the grand mean maybe more drastic. [Ref. 8: p. 123]

TABLE XV
FORMULAS

Let $N(t)$ = inventory at the beginning of fiscal year t (t = 1, ..., T)

Let y = number of attritions at any time during year t

Let n = maximum (y, $0.5(N(t) + N(t-1))$)

MAXIMUM LIKELIHOOD

$$m = \frac{y}{n}$$

MINIMAX

$$P = \frac{1}{(1+\sqrt{n})}\left(\frac{y}{\sqrt{n}} + \frac{1}{2}\right)$$

JAMES-STEIN

The following steps are utilized to generate James-Stein attrition rates.

STEP 1: Use a variance stabalizing transform (Freeman - Tukey).

$$x = \frac{(n+.5)^{.5}}{2}\left[Sin^{-1}\left(\frac{2y}{(n+1)} - 1\right) + Sin^{-1}\left(\frac{2(y+1)}{(n+1)} - 1\right)\right]$$

STEP 2: Form the cell means, the grand mean, SSB, and SSE

Let $\dot{X}$ = cell means

Let $\bar{X}$ = grand mean

Let K = number of treatment cells

STEP 3: Compute the set of James-Stein estimators in the transformed scale

$$C = 1 - \frac{T(K-3)}{K(T-1)+2}\,\frac{SSE}{SSB}$$

$$p = \bar{X} + C(\dot{X} - \bar{X})$$

STEP 4: Invert the transform to produce the attrition rates "r"

$$r = \frac{1}{2}\left[1 + Sin\left(\frac{p}{\sqrt{n+.5}}\right)\right]$$

Which set of values, "m" or "p", is the better indicator of the attrition rate? In order to answer that question one would have to know the true attrition rate. This true average shall be designated with "TH" in the main text of this paper and by the Greek letter theta in the Appendices. It is the probability that an officer would leave any time and is actually an unknowable quantity. Although "TH" is unobservable, an approximation can be made utilizing historical data.

One method of evaluating the two estimates, i.e., the James-Stein estimator and the observed attrition rate m, is by simply counting their closeness to "TH". Consider the closer the estimator is to "TH" as a success and the farther from "TH" as a failure. The number of successes of one estimator could then be compared to the number of successes of the other estimator.

A more quantitative method of comparing the two estimates is through the total squared error estimation. This is measured by first determining the actual error of each prediction , given by (TH-m) and (TH-p) for each cell. Each of these quantities is then squared and the squared values are added up. The observed averages "m" would have a total squared error denoted by Em, whereas the squared error of the James-Stein estimators would be denoted by Ejs. Comparing Em and Ejs by the ratio R=Em/Ejs, then the James-Stein method is "R" times as accurate as the observed attrition rate "m".

Employing the ideas borrowed from statistical decision theory, the estimators can be compared through a risk function. The risk function of a decision rule is the expected loss (over the sample) incurred using the decision rule [Ref. 9: p. 8]. The above techniques are similar to the squared-error loss method [Ref. 9: p. 9].

Therefore, the risk function is the expected value of the squared error for every possible value of "TH". Stein's theorem is concerned with estimation of several unknown means [Ref. 7]. No relation between the means need be assumed and indeed are assumed to be independent of one another. The historical development of statistical theory from Gauss through decision theory argues that the average is an admissible estimator as long as there is just one or two means to be estimated. An estimator for a parameter is admissible if, according to a specified criterion, there exists no other estimator that is better than it for all possible values of the parameter. Stein in collaboration with James showed that when the number of means exceeds two that estimating each of the means by their own average is an inadmissible procedure. No matter what the values of true mean, there are estimation rules with smaller total risk [Ref. 6].

The risk function for the James-Stein estimator is less than the risk function for the sample means irrespective what the true values of the means "TH" happen to be. The reduction of risk can be substantial, particularly when the number of means is relatively large. The risk of the James-Stein estimator is smallest when all the true means "TH" are equal. As the true means increase in variation from one another the risk of the estimator increases, approaching the value of the observed averages but never quite equaling it. The James-Stein estimator does substantially better than the averages only if the true means lie near each other. The James-Stein estimator does at least marginally better no matter what the true means are [Ref. 8: p. 124].

C. OTHER JAMES-STEIN MODELS

The model for the James-Stein estimator used thus far shrinks the observed averages to the grand mean. This is not the only possible procedure. Other models for the estimator dispense with the grand mean entirely. The observed attrition rates do not depend on a choice of origin. Before Stein discovered his method it was generally accepted that such "invariant" estimators must be preferable to those whose predictions change with each choice of an origin. If the origin or zero is chosen as the grand mean then the terms containing the grand mean would be removed from the equation. In this case the James-Stein estimator would be p=c(m). The estimation process is now complicated by the fact that the shrinking factor "c" would be different for each cell. The shrinking factor is dependent on the standard deviation of the sample attrition rates. A large standard deviation implies a high degree of randomness or uncertainty in the sample data. If the sample attrition is large, it can be attributed to random fluctuations rather than to an actual large value of the true mean "TH". Thus, application of a small shrinking factor would reduce the value. [Ref. 7: pp. 123-4]

There is one purpose for which the measured sample attrition rate may well be superior to the James-Stein estimator: when a single cell is considered in isolation. The James-Stein method gives better estimates for a majority of cells and it reduces the total error of estimation for the sum of all the cells. Estimating the true mean for an isolated cell by Stein's method creates serious errors when that mean has an atypical value. The reduction of the risk is more predominant in a homogeneous setting. The inclusion of an atypical or nonhomogeneous mean in the estimation process would increase the risk.

1. Bayesian versus James-Stein

The formula for the James-Stein is similar to the Bayesian equation of Z=m+C(y-m). Here "y" is the sample mean, "m" is the mean of the prior distribution. The shrinking factor "C" is different in that it depends on the standard deviation of the prior distribution. [Ref. 9: pp. 99-110]

The James-Stein procedure, however, has one important advantage over the pure Bayesian method. The James-Stein estimator can be employed without knowledge of the prior distribution. The James-Stein estimator can be regarded as an empirical Bayes rule. The empirical Bayes approach uses historical data to estimate the prior distribution [Ref. 9: p. 117].

2. Modifications to the James-Stein Process

The James-Stein estimation process can be modified using the Efron-Morris limited translation version of the estimator [Ref. 10]. This modification ensures that the estimator of "TH" is not shifted so far from the sample mean that the estimator is inconsistent with the sample mean. The modification is a compromise between the James-Stein estimator and the Maximum Likelihood Estimator which has good individual properties. The compromise would follow the James-Stein rule as closely as possible subject to a fixed constraint on how far the estimator is allowed to deviate from the MLE. Such "limited translation estimators" were discussed in a Bayesian framework in the Efron-Morris article [Ref. 11]. They showed that it is possible to considerably reduce the maximum possible risk for any component while reducing the ensemble savings of the James-Stein estimator by as little as 5 or 10 percent [Ref. 10].

D. VALIDATION TESTS

A validation test was conducted to evaluate the efficiency of the James-Stein shrinkage estimator. The test was conducted as follows:

1. Select a grade within an Occupational Group to examine. The resultant desired data array will be three dimensional (years, LOS, MOS).
2. Let "i" stand for LOS, then i=0,...,30.
3. Let "j" stand for MOS, (then values of j's depend on which MOS group is being analized).
4. Let Dij = Incidence Matrix of nonstructural zeroes, which is the same for all years. Dij = 1, if cell is member of the feasible region; Dij = 0, otherwise.
5. Let K = number of feasible cells, i.e., sum of all Dij.
6. Let Yij = number of leavers in cell (i,j).
7. Let t = 1,...,T; where T = number of years of data used to create the estimator.
8. Let Nij = Inventory in cell (i,j) = Max((N(t)+N(t+1))/2,Y(t)).

The validation procedures used t=1...4 to compute the empirical estimates and used t=5,6 for validation purposes. Three estimation methods were employed: James-Stein, Minimax, and Maximum Likelihood.

1. Preliminary Steps:

The following steps were utilized to prepare the data for the validation procedures. These steps included the transformation of the given data of leavers and computed inventory according to the variance stabilizing arc sine transformation listed in Table XV. See Appendix D, Figure D.18 for the APL listings of the following functions.

- Let ISij(t) = Yij(t) BINPREP Nij(t), where "BINPREP" prepares the Freeman-Tukey version of the arc sine transformation for binomial data.
- Let T = 4, years of data desired.
- Let Jij = Dij JAMES ISij(t), where "JAMES" returns a James-Stein estimator for the means of the cells in the last two dimensions of "IS" while being screened by the incidence matrix "D" of nonstructural zeroes.
- Let Pij = Jij BINCONV Nij(t), where "BINCONV" inverts the arc sine transformation used in preliminary step one.

2. <u>Validation Procedures</u>:

The following procedures were utilized to validate the effectiveness of the the James-Stein estimation process.

- Let IS*ij = ISij(T+1), which is assumed to be distributed Normal(Jij,unknown variance).
- Let NOR = IS*ij - Jij, where Jij was derived in preliminary step three.
- Use the distribution fitting capability of GRAFSTAT or any other comparable graphical display package, to compare NOR to the Normal distribution. Desired output is a comparison histogram of the data to a fitted Normal density, compare sample distribution to fitted Normal distribution, normal probability Q-Q plot with data, and possibly the survivability curve fit. These plots can all be plotted on one sheet of output.
- Compute test statistics comparing data fitted to a Normal distribution.
- Compute the figure of merit from the test statistics using the sum of the square of the sample mean and the square of the sample standard deviation.

3. Minimax Estimator

The following procedures were utilized to compare the James-Stein estimator to the Minimax estimator. See Table XV for the formula for the Minimax estimator. An simple example of the possible values resulting from the use of this formula is provided in Table XVI, to illustrate the shrinkage characteristics inherent in its use for the purpose of projecting rates which have extreme empirical values.

TABLE XVI

MINIMAX EXAMPLE VALUES

LEAVERS	INVENTORY	MINIMAX ESTIMATE
0	0	.50
0	1	.25
1	1	.75
0	4	.17
1	4	.33

See Appendix D, Figure D.19 for the APL listings of this function.

- Let PMMij = Yij(t) MINMAX Nij(t), where "MINMAX" returns the MINMAX estimates for the binomial.
- Let ISM = Nij ARCSIN PMMij, where "ARCSIN" returns the inverse sine transformation for use when the success probability PMMij is estimated directly (e.g., by MINMAX).
- Let NORMM = IS* - ISM, where IS* is from validation procedures step one.

- Use the distribution fitting capability of GRAFSTAT or any other comparable graphical display package, to compare NORMM to the Normal distribution similar to the validation procedure three.
- Compute test statistics comparing data fitted to a Normal distribution.
- Compute the figure of merit from the test statistics using the sum of the square of the sample mean and the square of the sample standard deviation.

4. Maximum Likelihood

The natural empirical estimator without shrinkage is the Maximum Likelihood estimator. Comparison of the James-Stein process to the Maximum Likelihood process will illustrate the efficiency of the James-Stein estimation process.

- Let Y = sum of Yij(t) for t=1 to T. Where T = 4.
- Let N = sum of Nij(t) for t=1 to T. Where T = 4.
- Let ISAij = Y BINPREP N.
- Let NORA = IS* - ISAij, where IS* is from the validation procedures step one.
- Use the distribution fitting capability of GRAFSTAT or any other comparable graphical display package, to compare NORA to the Normal distribution similar to the validation procedure three.
- Compute test statistics comparing data fitted to a Normal distribution.
- Compute the figure of merit from the test statistics using the sum of the square of the sample mean and the square of the sample standard deviation.

E. DATA ANALYSIS

A data analysis was conducted utilizing the three comparison mentioned above.

1. The James-Stein Estimator

Four years of data (1977-1980) was used to construct the James-Stein estimator. The leavers and inventory were transformed with the Freeman-Tukey transformation. The projected year's (1981-1983) leavers and inventory were constructed and also transformed with the Freeman-Tukey transformation. The difference between the James-Stein estimator and the projected years data was plotted to compare the fit to a Normal distribution. The figure of merit (FOM) was computed from the inputs taken from the statistical table.

$$\text{FOM} = (\text{squared mean}) + (\text{variance}).$$

Table XVII illustrates the results of the computation of the James-Stein, Minimax, and Maximum Likelihood FOM's. The data organization format of the table is provided for ease of interpretation of the table. The table demonstates the calculation of the FOM for the years 1981, 1982, and 1983. The separate cases evaluated are identified by rows which represent the grades of First Lieutenant and Lieutenant Colonel in each of the four MOS groups utilized in this project.

2. The Minimax Estimator

The leavers and inventory of the four years (1977-1980) were used to construct a minimax estimator. See Table XV for the formula utilized to derive this estimator. The result was compared to the fit of the Normal distribution in a similar manner as discussed above. The FOM was

computed as shown above and further illustrated in Table XVII.

3. Maximum Likelihood

The last comparison was using the leavers and inventory of the four years (1977-1980) and only using the Freeman-Tukey transformation. The result is the Maximum Likelehood estimator in the transformed space. The algorithm is the same as James-Stein estimator described above, only the James-Stein calculation is by-passed. The projected years data was used to compare the resultant difference with the fit to the Normal. The FOM was computed as shown above and further illustrated in Table XVII.

4. Figure of Merit

The figure of merit is representative of the efficiency of the estimator. Since the FOM is computed in the transformed and Normalized scale, the smaller the FOM, the better the estimate. The ratios illustrated in columns 3 and 5 of Table XVII illustrate (in most cases) how many times larger the other estimating technique's FOM is when compared to the James-Stein estimator technique. One can note that in one case the use of the James-Stein estimator technique resulted in an improvement of as much as 30 times over the Maximum Likelihood process and 22 times over the Minimax process. These results are typical of other cases executed by the author, but not shown for redundancy sake.

F. ADDITIONAL COMPARISONS

The above risk comparisons were all performed in the transformed (Freeman-Tukey) space. Since this distorts the scale it seems wise to perform some additional comparisons (sum of squares error) in the original (untransformed)

TABLE XVII

FIGURE OF MERIT COMPARISONS

*********** DATA ORGANIZATION FORMAT ***********

*** COLUMN ***	*** REMARK ***
JAMES	FOM of data using James-Stein
MAXLIKE	FOM of Maximum Likelihood
MAXLIKE/JAMES	Ratio of Maxlike divided by James
MINIMAX	FOM of data using Minimax
MINIMAX/JAMES	Ratio of Minimax divided by James

*** ROW ***	*** REMARK ***
1	1ST LT in Aviation
2	1ST LT in Combat Service Support
3	1ST LT in Combat Support
4	1ST LT in Ground Combat
5	LTCOL in Aviation
6	LTCOL in Combat Service Support
7	LTCOL in Combat Support
8	LTCOL in Ground Combat

**** THESE ARE THE FIGURES OF MERIT FOR 1981 ****

ROW	JAMES	MAXLIKE	MAXLIKE/ JAMES	MINIMAX	MINIMAX/ JAMES
1	1.69	51.47	30.46	37.68	22.30
2	0.71	1.60	2.25	0.78	1.10
3	1.20	3.69	3.08	2.10	1.75
4	1.19	5.47	4.59	3.81	3.20
5	3.05	13.64	4.47	8.92	2.92
6	0.22	0.42	1.91	0.23	1.05
7	0.31	2.34	7.55	1.16	3.74
8	0.79	6.81	8.62	3.73	4.72

**** THESE ARE THE FIGURES OF MERIT FOR 1982 ****

ROW	JAMES	MAXLIKE	MAXLIKE/ JAMES	MINIMAX	MINIMAX/ JAMES
1	4.37	47.05	10.77	34.29	7.85
2	0.75	1.55	2.07	0.76	1.01
3	1.53	3.71	2.42	2.19	1.43
4	1.89	5.84	3.09	4.21	2.23
5	5.80	17.28	2.98	13.08	2.26
6	0.23	0.54	2.35	0.33	1.43
7	0.59	2.60	4.41	1.43	2.42
8	1.13	8.34	7.38	4.92	4.35

**** THESE ARE THE FIGURES OF MERIT FOR 1983 ****

ROW	JAMES	MAXLIKE	MAXLIKE/ JAMES	MINIMAX	MINIMAX/ JAMES
1	4.64	43.16	9.28	30.65	6.63
2	0.91	1.64	1.80	0.86	0.95
3	1.17	3.67	3.14	2.02	1.73
4	2.85	5.67	1.99	4.21	1.48
5	5.49	11.93	2.17	8.44	1.54
6	0.26	0.56	2.15	0.34	1.31
7	0.62	2.76	4.45	1.53	2.47
8	1.57	8.55	5.45	5.16	3.29

scales. Accordingly the deviations of the empirical attrition rates in the validation time frame from the estimates (both James-Stein and Maximum Likelihood) are computed and compared. Specifics follow:

The preliminary steps are as follows:

- Actual rates: empirical attrition rates are the central attrition rates for the projected period.
- James-Stein Projected rates: projected attrition rates utilizing the James-Stein estimation process previously defined.
- Let, MLE= Maximum Likelihood Estimates, the historical empirical attrition rates are assumed to remain the same throughout the projected period.
- Let, SEP= The squared differences between the James-Stein projected attrition rates and the actual rates.
- Let, SEM= The squared differences between the MLE and the actual rates.
- Let, SSE(MLE)= The sum of squared differences of SEM.
- Let, SSE(P) = The sum of squared differences of SEP.
- Let, SEP<SEM be the number of cells for which SEP was less than SEM. (For convenience of column heading, inequalities are allowed to represent numbers.)
- Let, SEM<SEP be the number of cells for which SEM was less than SEP.
- MOS Groups evaluated were: CS - Combat Support Group; GC - Ground Combat
- Ranks evaluated were: LT - First Lieutenant; LTCOL - Lieutenant Colonel

1. Subtest 1

First it was assumed that all cells in the feasible region were eligible for forecasting the projected attrition rates. The results of subtest 1 are illustrated in Table

XVIII. The sum of squared differences for the MLE and the James-Stein projected 'rates for the first and last groups are competitive. One has to keep in mind that the Jame-Stein estimation process will result in a value greater than zero given a cell with an original value of zero. Therefore, when all the cells in the feasible region are considered eligible for projection, the empirical values of zero result in a value closer to the grand mean. The value of the sums of the differences between the James-Stein estimation process and the actual rates can be understandably larger than the sums of the differences between the MLE and the actual rates when there are numerous zero cells originally in the feasible region.

TABLE XVIII

SUBTEST 1

Group	Rank	Feasible Cells	SEP<SEM	SEM<SEP	SSE(MLE)	SSE(P)
CS	LT	57	15	42	1.52	1.91
GC	LT	57	22	35	0.33	3.99
CS	LTCOL	48	12	36	1.29	2.93
GC	LTCOL	48	21	27	2.05	0.66

2. Subtest 2

A second comparison assumed a conditioning that considered only the cells in which there was an actual attrition rate greater than zero. In other words, what was the projected attrition rate given an actual loss occured? This created a subfeasible region which was smaller than the

whole feasible region. The results of subtest 2 are illustrated in Table XIX. The sum of squared differences for the James-Stein projected rates in this subtest are much lower than the MLE.

TABLE XIX

SUBTEST 2

Group	Rank	Sub-Feasible Cells	SEP<SEM	SEM<SEP	SSE(MLE)	SSE(P)
CS	LT	26	15	11	1.52	0.27
GC	LT	29	22	7	0.33	0.10
CS	LTCOL	16	12	4	1.29	0.14
GC	LTCOL	28	21	7	2.05	0.12

VI. CONCLUSIONS AND RECOMMENDATIONS

A. SUMMARY

The purpose of this study was to demonstrate the application of the James-Stein and other shrinkage type parameter estimation schemes for the goal of generating manpower loss rates within the USMC officer force structure. This thesis contains comparisons of performance of James-Stein, Minimax, and Maximum Likelihood estimations of Marine Corps officer attrition rates for select cell aggregates. The very large number of cells within the USMC officer force structure leads to the condition that empirical attrition rates are unstable. This problem is compounded by the fact that the cell probabilities are small. Most rates are less than 10% and virtually none are greater than 20% at disaggregated levels. Further difficulties are present because some of the inventory cells are empty for structural reason while others are empty by chance. Therefore, the small inventory cells draw especial attention. It has been illustrated within this project that improvement can be attained by application of the James-Stein and Minimax shrinkage methods rather than the more natural Maximum Likelihood estimation process. It is important to note that the James-Stein and Minimax shrinkage schemes seem to compete for appropriate attrition rate generation.

B. CONCLUSIONS

The shrinkage schemes employed herein offer powerful and useful methods for generating attrition rates. This employment should lead to much lower costs that accrue from errors in manpower planning. However, the particular way to explicit them is yet to be determined.

The application of the James-Stein shrinkage technique has exposed a number of problem areas which could be expanded.

The aggregation problem. Aggregation of cells (cross classified by grade, Military Occupational Specialty, length of service, etc) into sets whose attrition behavior is homogeneous and of low internal variability should lead to the most useful overall performance. One method, not attempted by the author, is to aggregate over MOS types in which the grade structure is similar. For example, some MOS's are restricted to warrant officers only, warrant officers and limited duty officers only, unrestricted regular/reserve officers only, etc. Another method would be to aggregate over certain length of service periods. For example, 1-4 years of service, 5-10 years of service, 10-20 years of service, and 21-30 years of service. Thus examining the attrition rates in relationship to breakpoints and longevity in the officer's career pattern.

The limited translation problem. The shrinkage estimators translate the rates toward the aggregate mean. Such translations can appear to be excessive in instances of extreme rates. This suggests that techniques that limit the translation could be profitably applied. Current literature contains limited translation methods which could be examined.

The yearly update problem. The estimated attrition rates are based upon data from several recent years. As each year produces a new set of experiences there is need for a policy to include new data in the estimation scheme and phase out the influence of the distant past.

The validation problem. The smallness of the rates (i.e., 0-20%) places the current project in a range which has not yet been treated successfully in the literature. For this reason it is especially important to apply cross validation procedures to score the efficacy of the methods.

The multinomial estimation problem. The number of attritions in a cell are placed into a number of (currently eight) disjoint categories. This produces a number of small sample size multinomial probability estimation problems. Methodologies for managing such problems have attracted attention in the recent literature. Their usefulness could be further examined in regard to the officer attrition rate estimation process

C. RECOMMENDATIONS

It is recommended that further studies pursue the problems identified above. It is suggested that a new data tape be provided to the Naval Postgraduate School which allows the distinction between restricted and unrestricted officers. Additionally, the data for fiscal years 1984 and 1985 needs to be provided in the near future. In order to provide continued service, a regularly scheduled yearly updated data set should be forwarded to the Naval Postgraduate School when available.

APPENDIX A

CENTRAL ATTRITION RATE GRAPHICS

This appendix contains graphical illustrations of the central attrition rate for USMC officers.

Figures A.1 to A.10 contain the central attrition rates for each Military Occupational Specialty distributed over the length of service period from 0 to 30. These figures illustrate the aggregate central attrition rate for all eight types of losses over the six years from 1977 to 1982. These figures demonstrate the breakpoints within each MOS over the distribution of length of service.

Figures A.11 to A.20 contain the central attrition rates for each Military Occupational Specialty distributed over the grades from Warrant Officer to Colonel (labelled 0...9 on the abscissa). These figures also illustrate the aggregate central attrition rate for all eight types of losses over the same six years.

Figures A.21 to A.28 contain the central attrition rates for each Military Occupational Specialty distributed over the years from 1977 to 1982. Each figure has five MOS graphs and are distinguishable by the type of line and symbol points for each separate MOS. These figures illustrate the aggregate central attrition rate for all eight types of losses over the six years from 1977 to 1982.

Figure A.29 contains the central attrition rates for the Aviation Group for the grades from Second Lieutenant through Lieutenant Colonel for the years 1977 to 1982. Figure A.30 contains the central attrition rates for the Combat Support Group for the grades from Second Lieutenant to Lieutenant Colonel for the years 1977 to 1982. Figure A.31 contains the central attrition rates for the Ground Combat Group for

the grades from Second Lieutenant to Lieutenant Colonel for the years 1977 to 1982. Figure &TOTMOS contains the central attrition rates for the Total MOS's for the grades from Second Lieutenant to Lieutenant Colonel for the years 1977 to 1982.

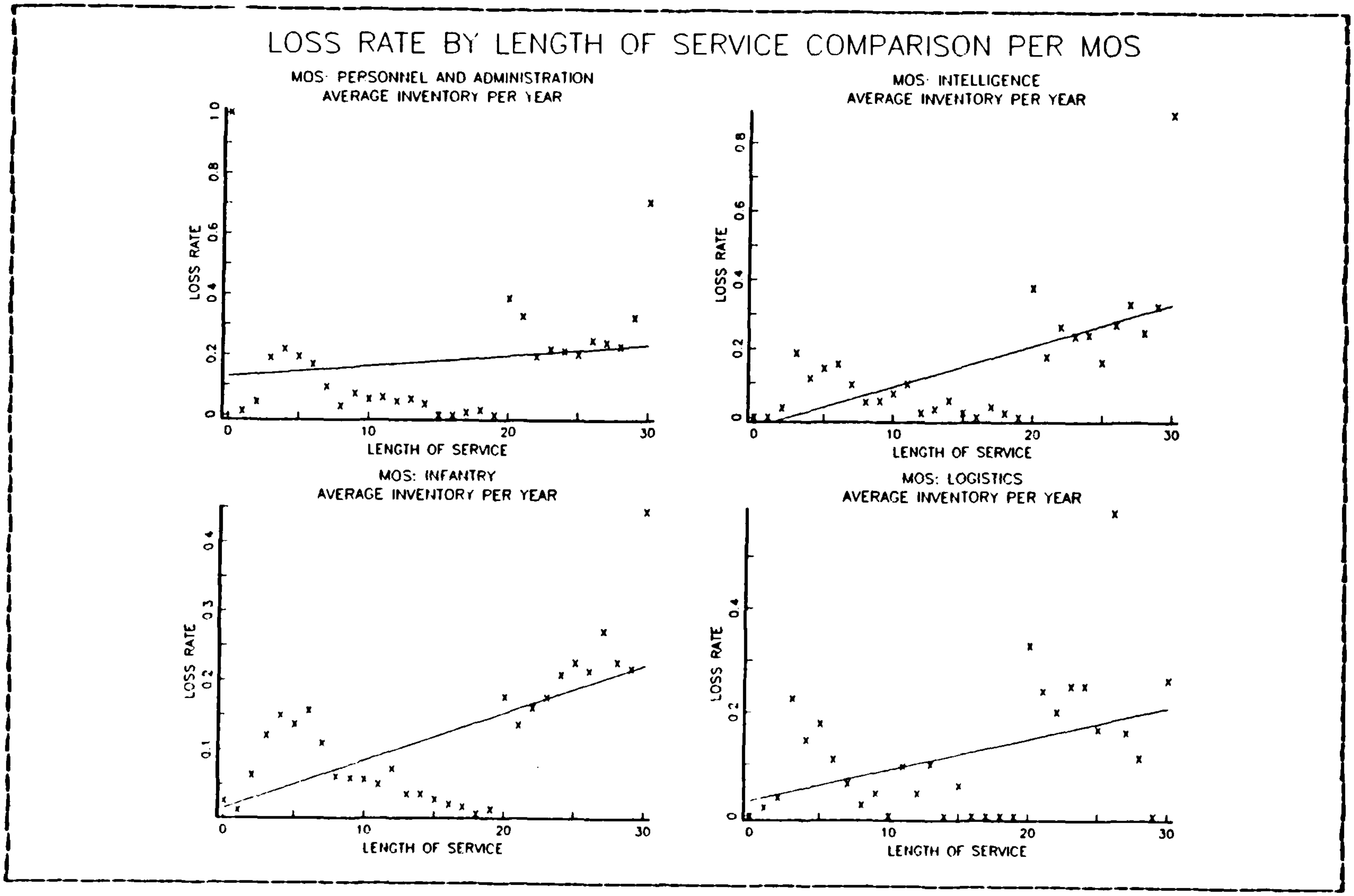


Figure A.1 Attrition Rates by LOS for MOS's 01 to 04.

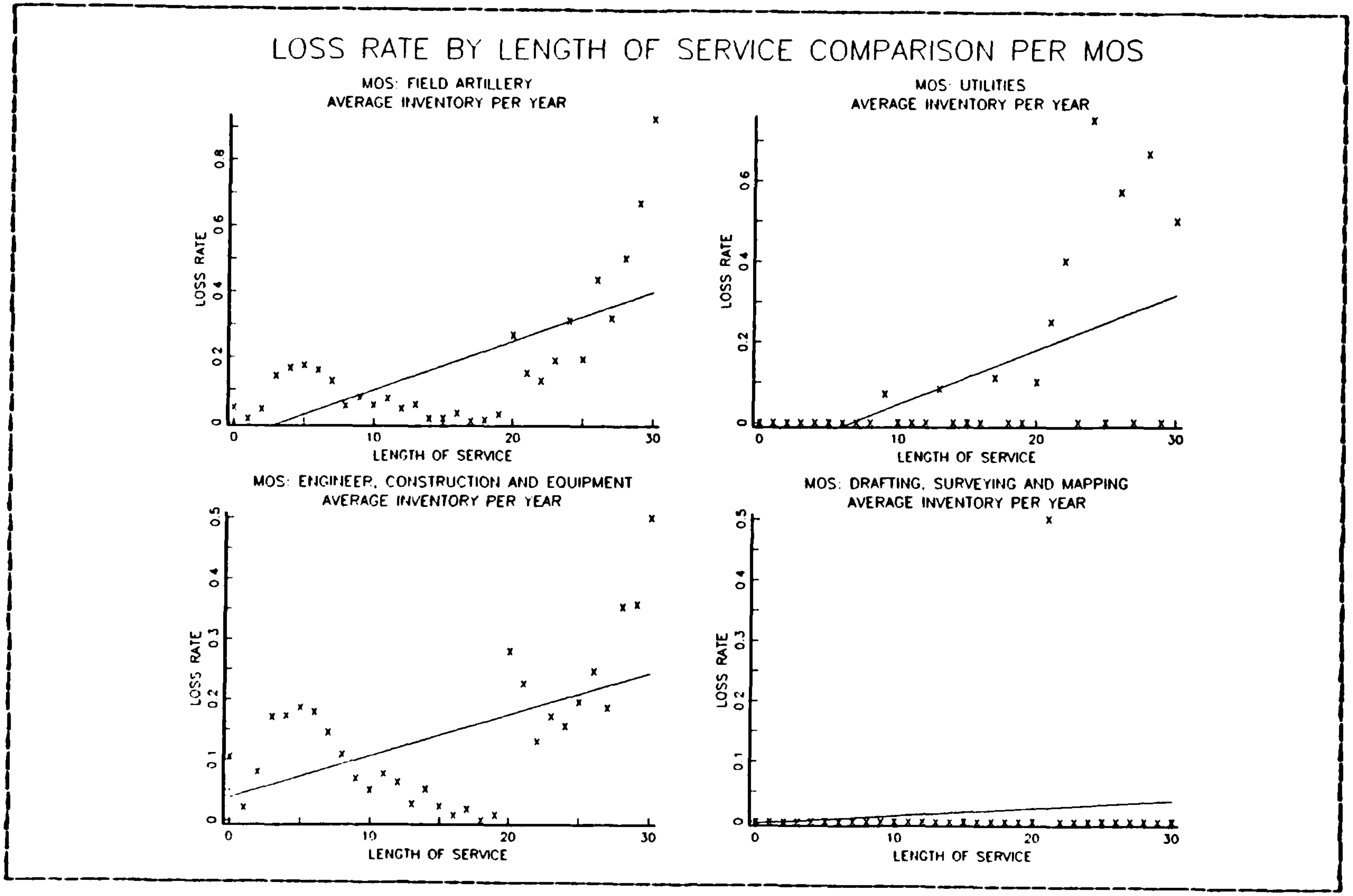


Figure A.2 Attrition Rates by LOS for MOS's 05 to 08.

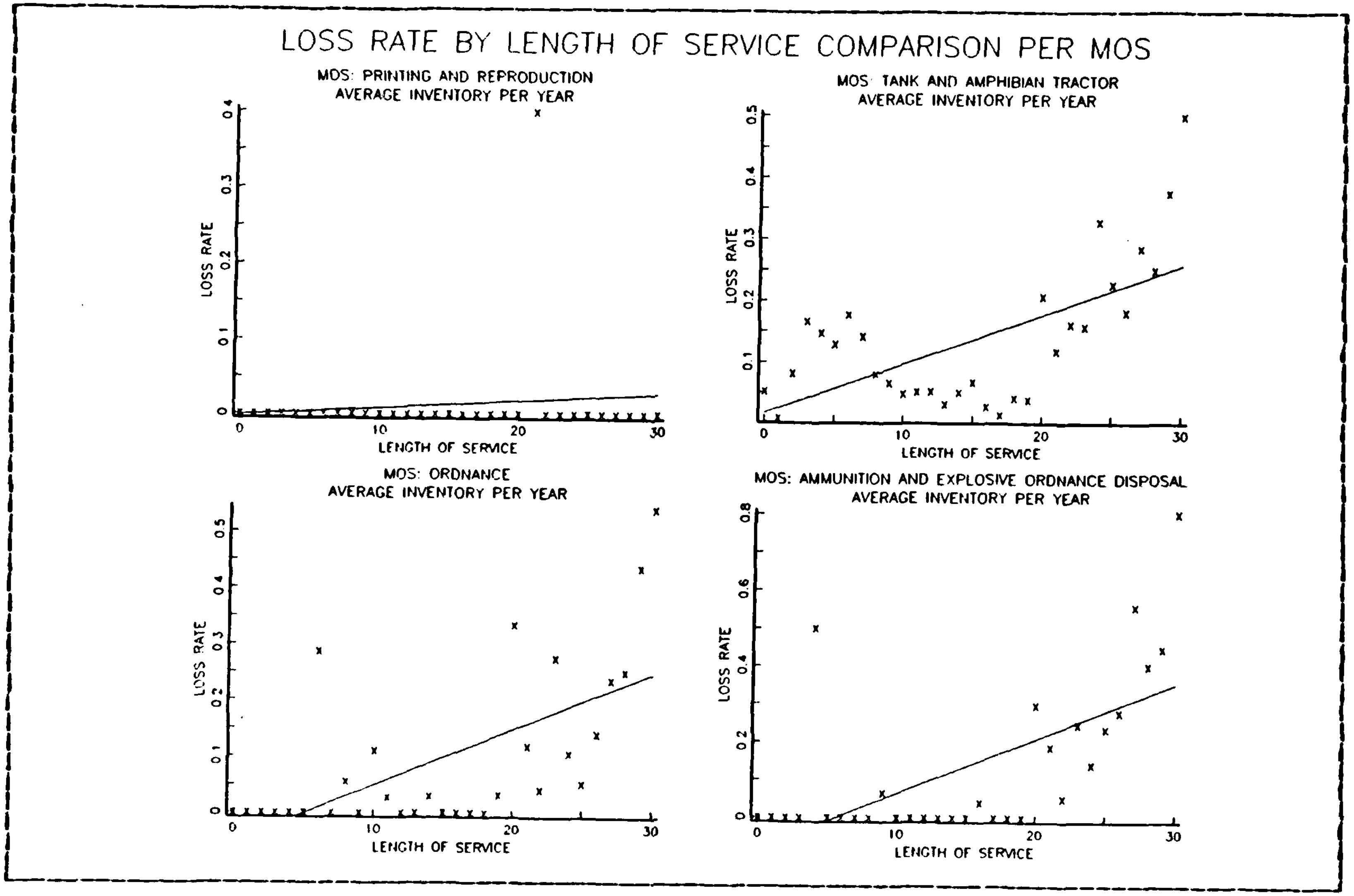


Figure A.3 Attrition Rates by LOS for MOS's 09 to 12.

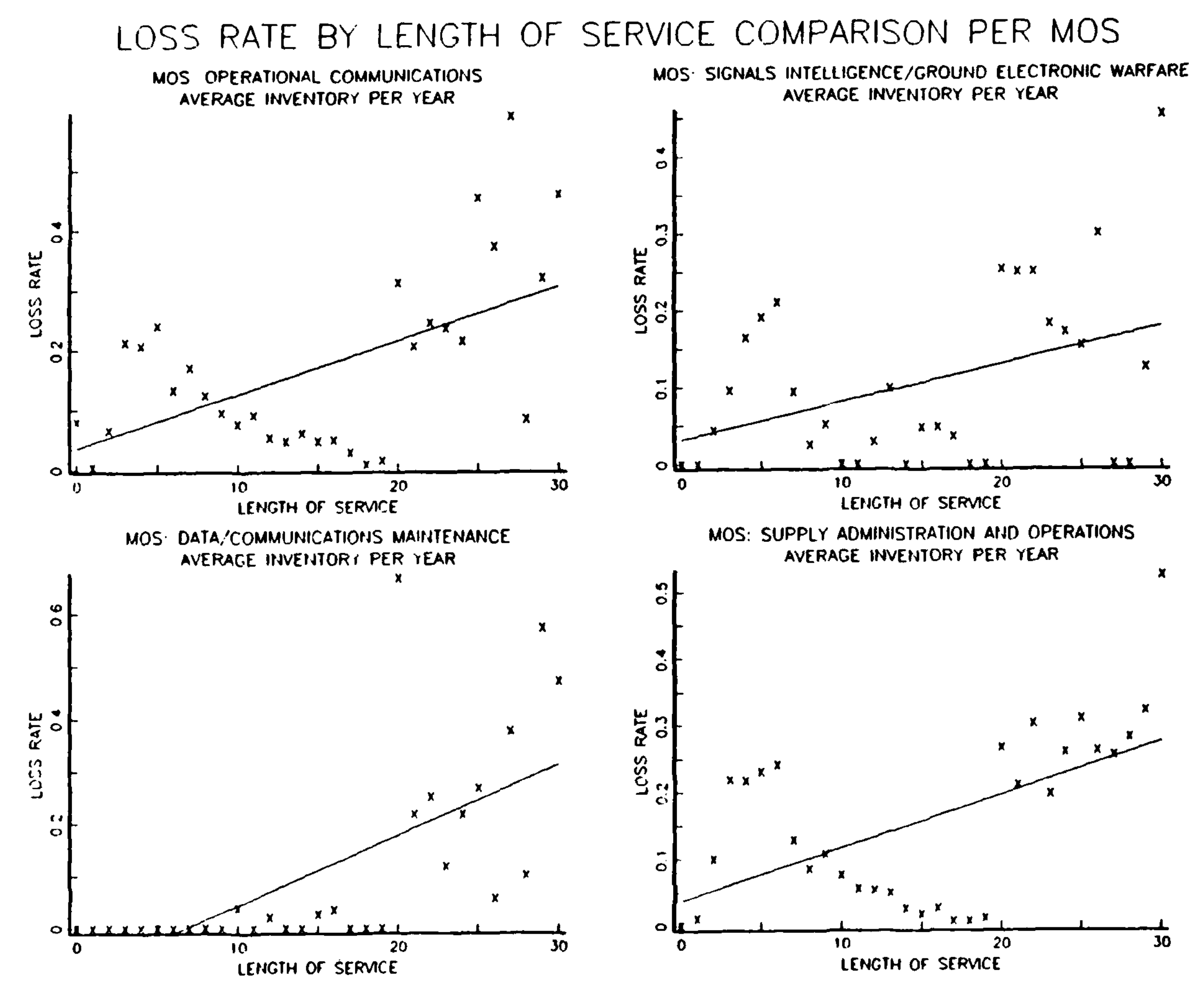


Figure A.4 Attrition Rates by LOS for MOS's 13 to 16.

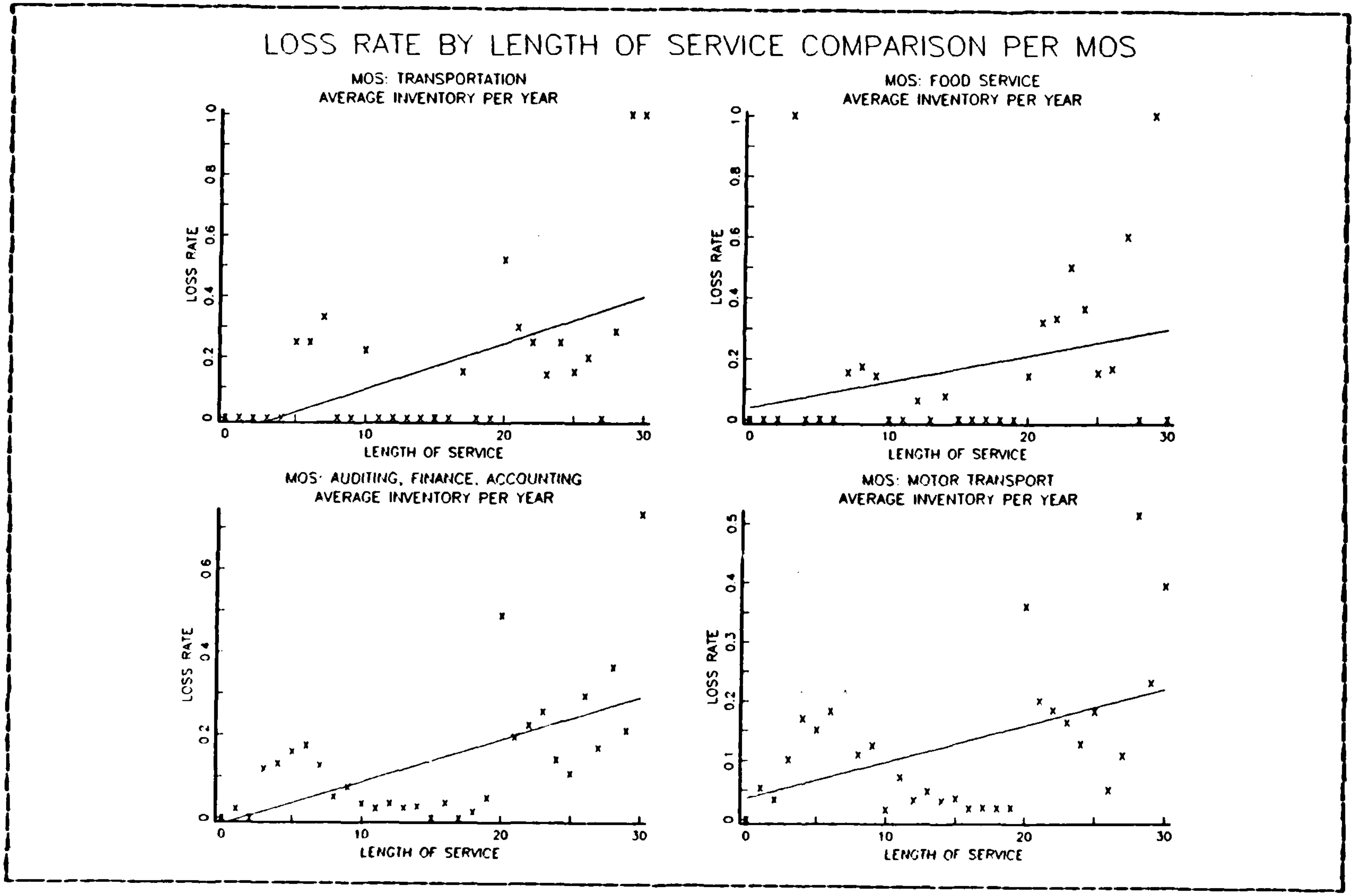


Figure A.5 Attrition Rates by LOS for MOS's 17 to 20.

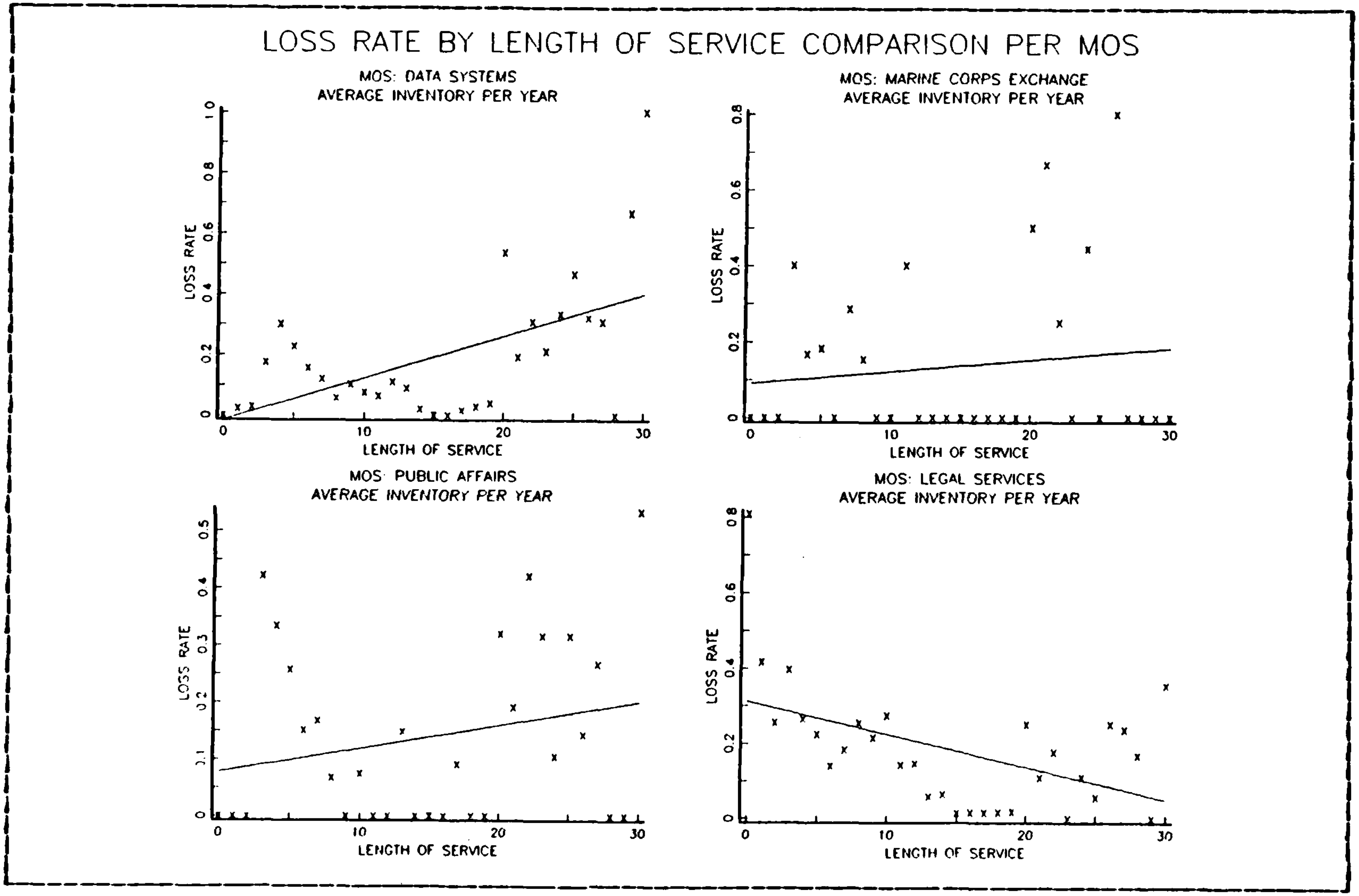


Figure A.6 Attrition Rates by LOS for MOS's 21 to 24.

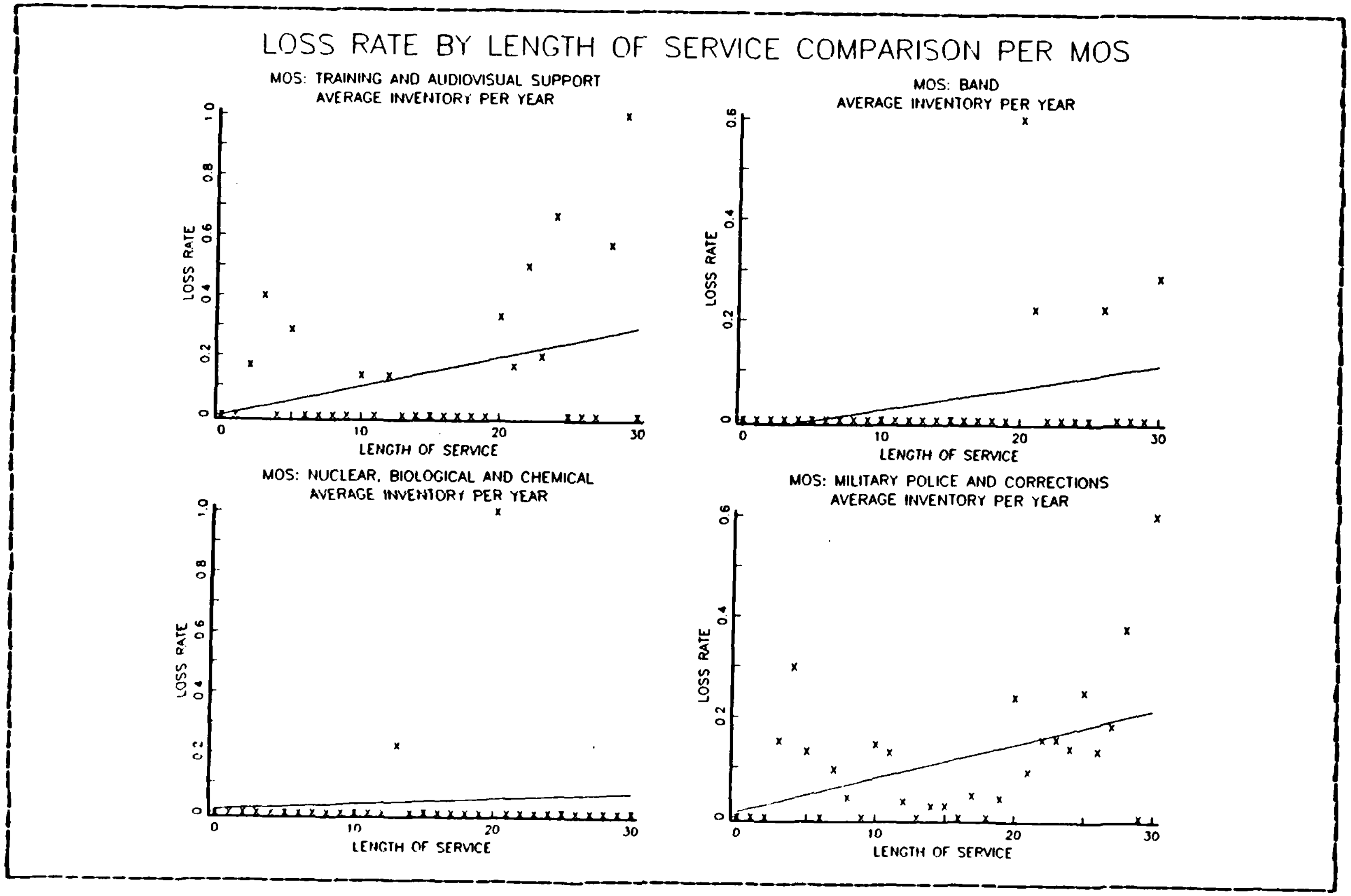


Figure A.7 Attrition Rates by LOS for MOS's 25 to 28.

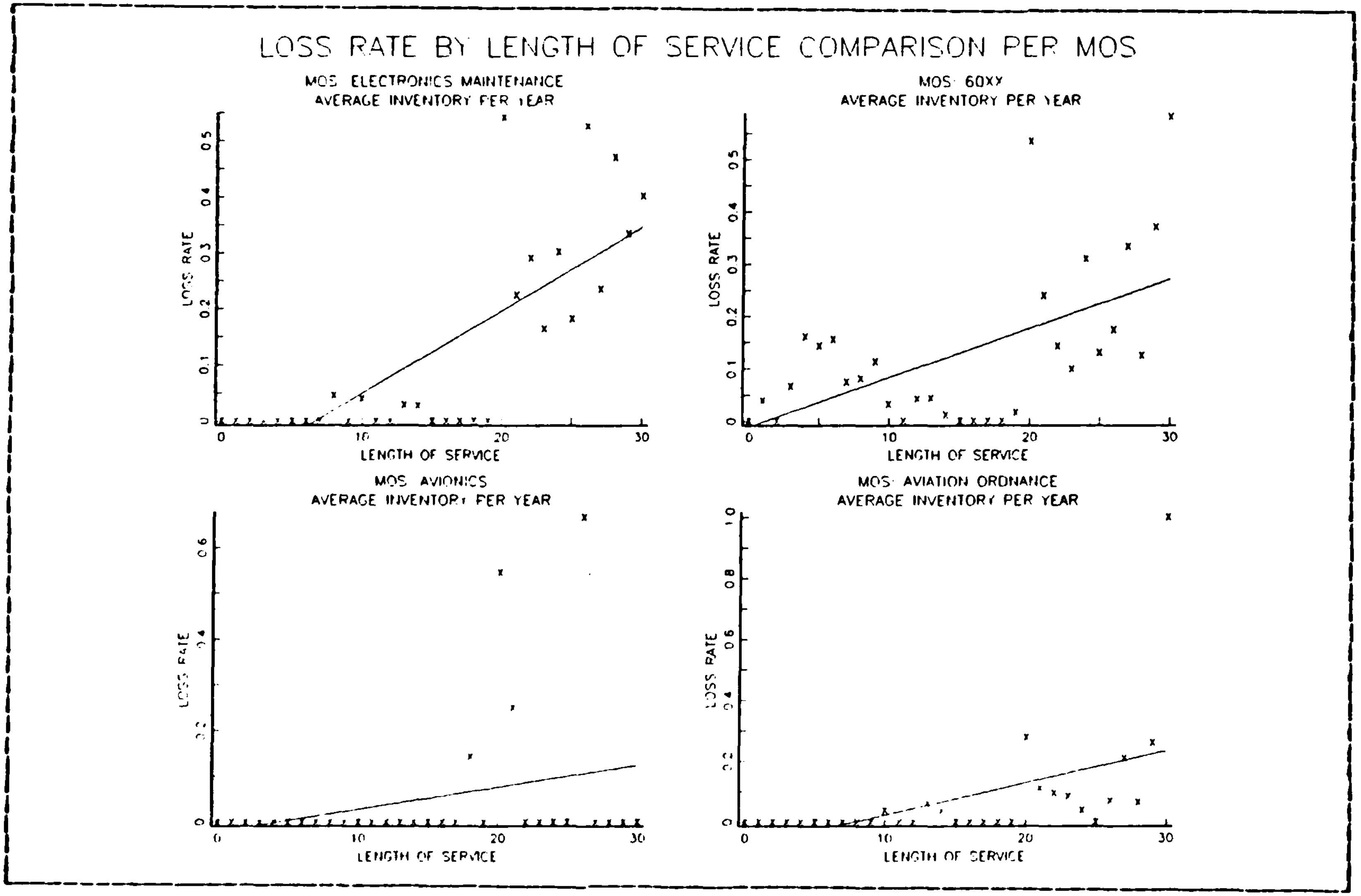


Figure A.8 Attrition Rates by LOS for MOS's 29 to 33.

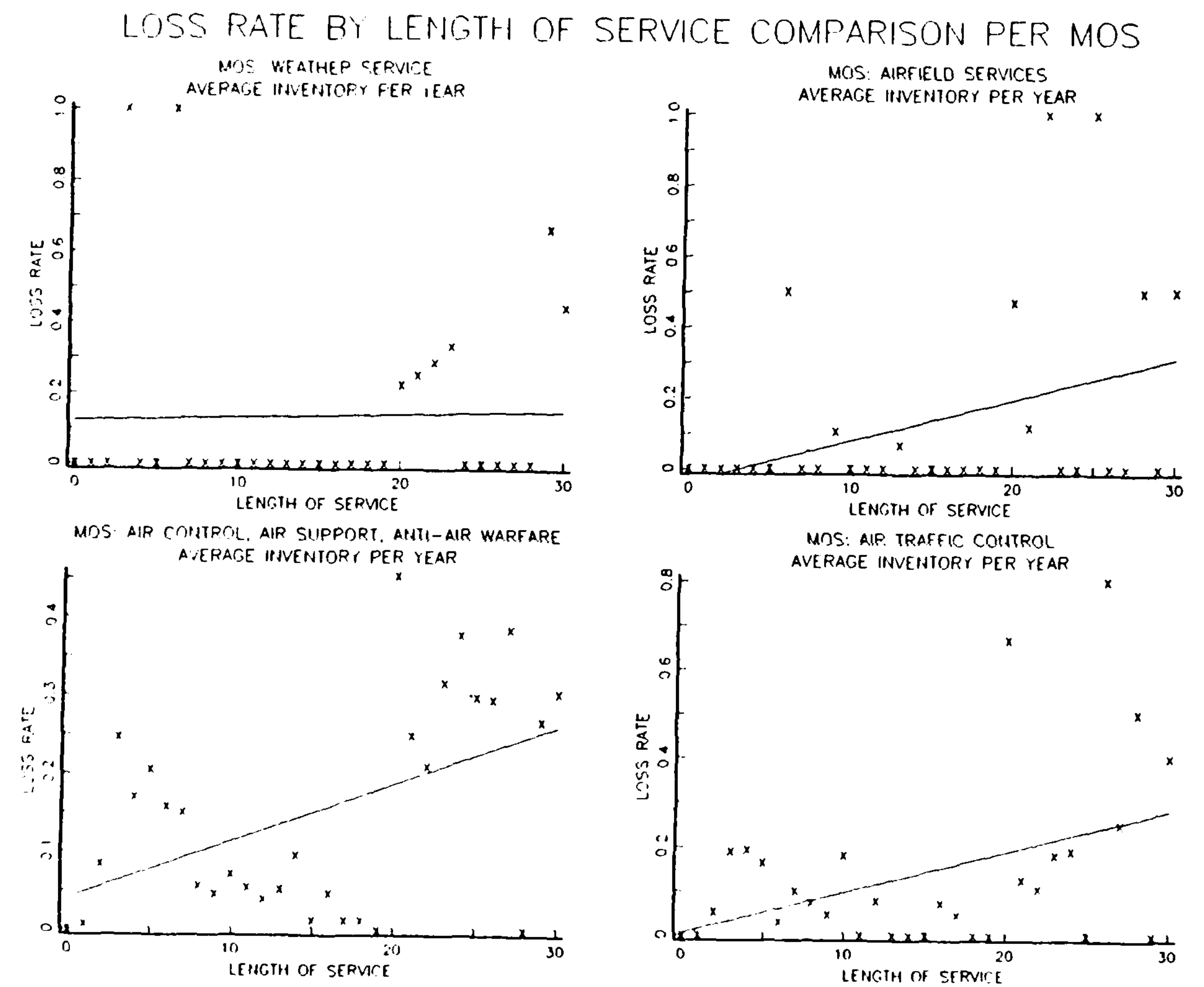


Figure A.9 Attrition Rates by LOS for MOS's 34 to 37.

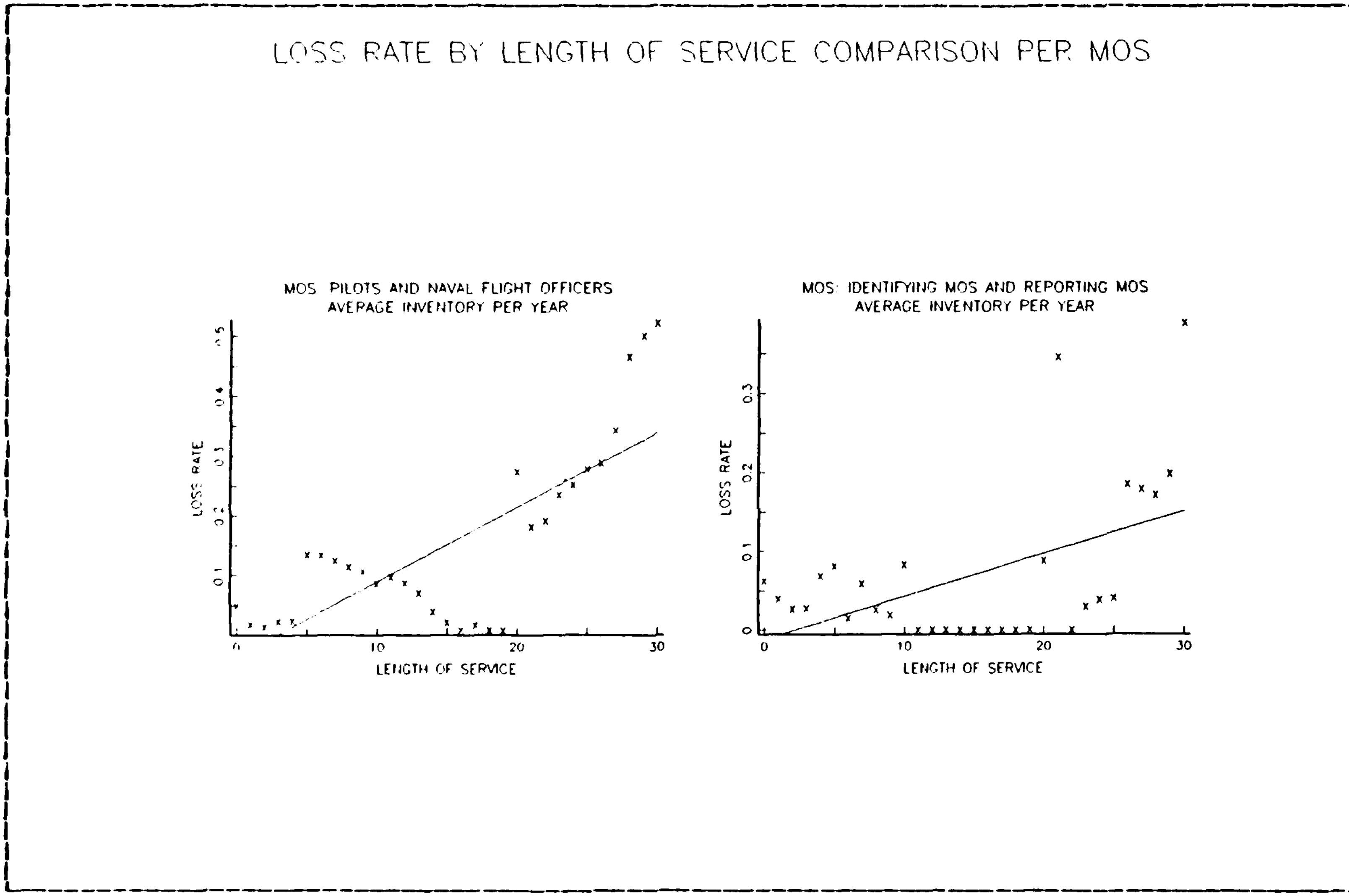


Figure A.10 Attrition Rates by LOS for MOS's 38 to 39.

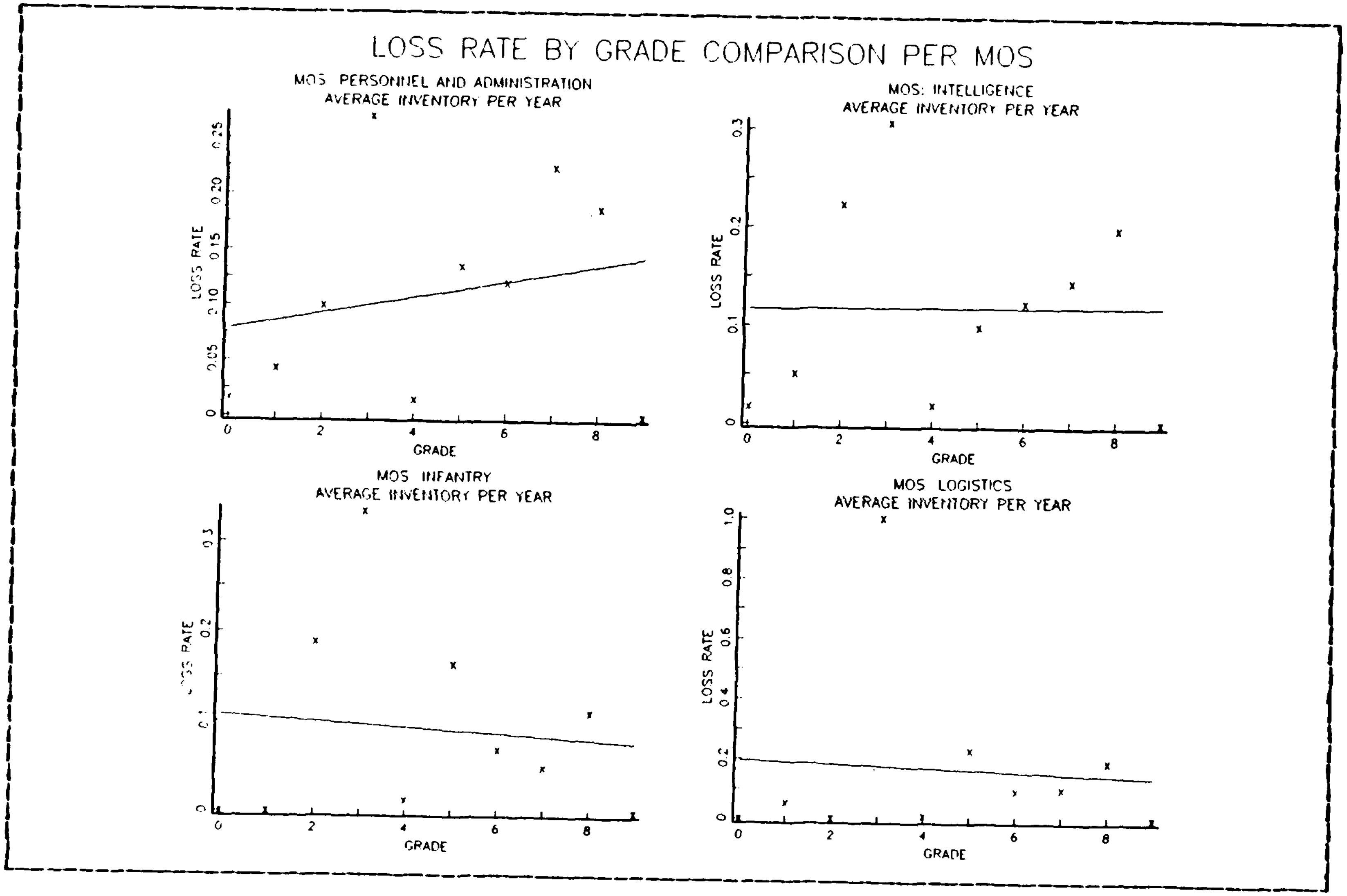


Figure A.11 Attrition Rates by Grade for MOS's 01 to 04.

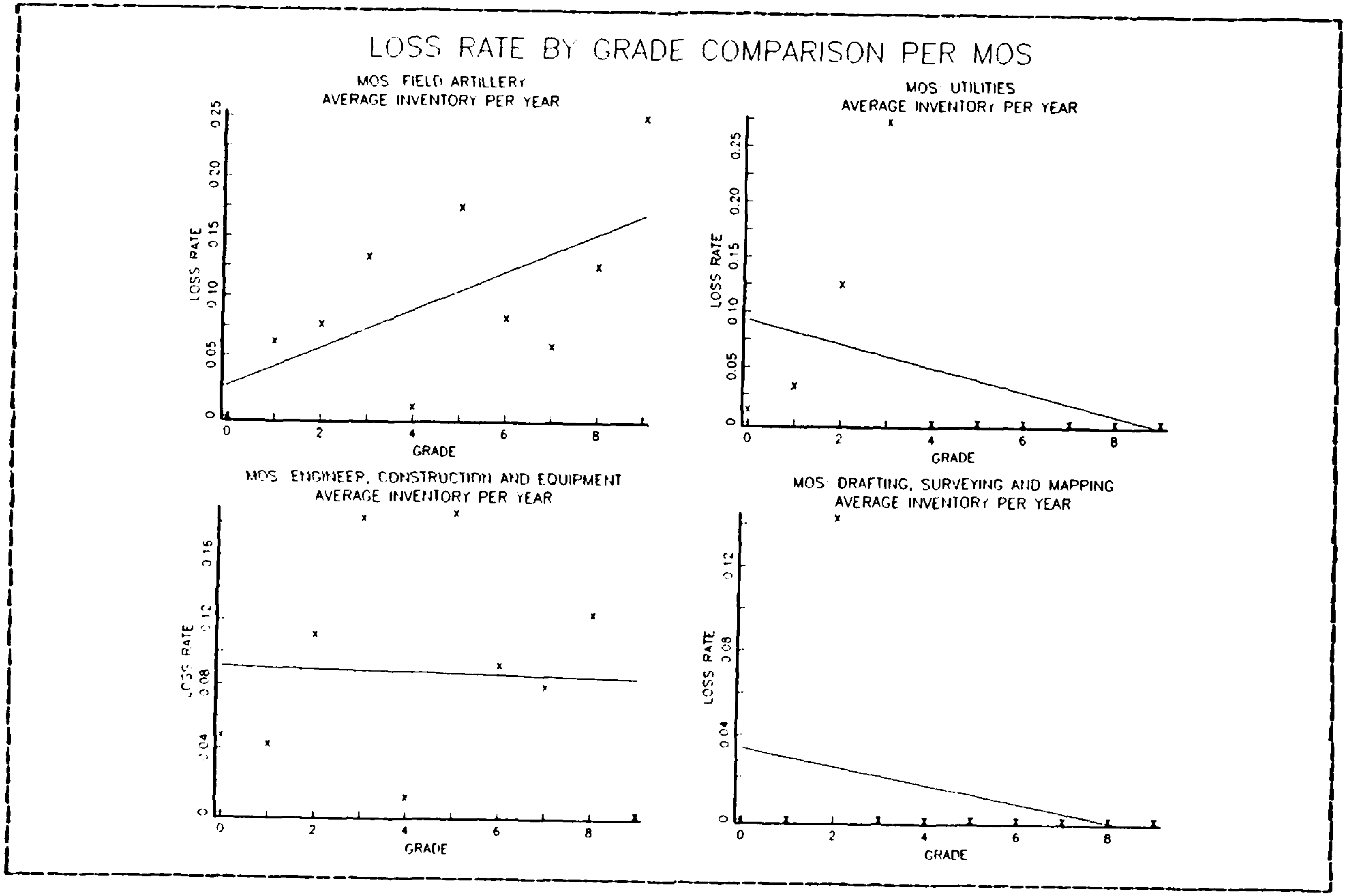


Figure A.12 Attrition Rates by Grade for MOS's 05 to 08.

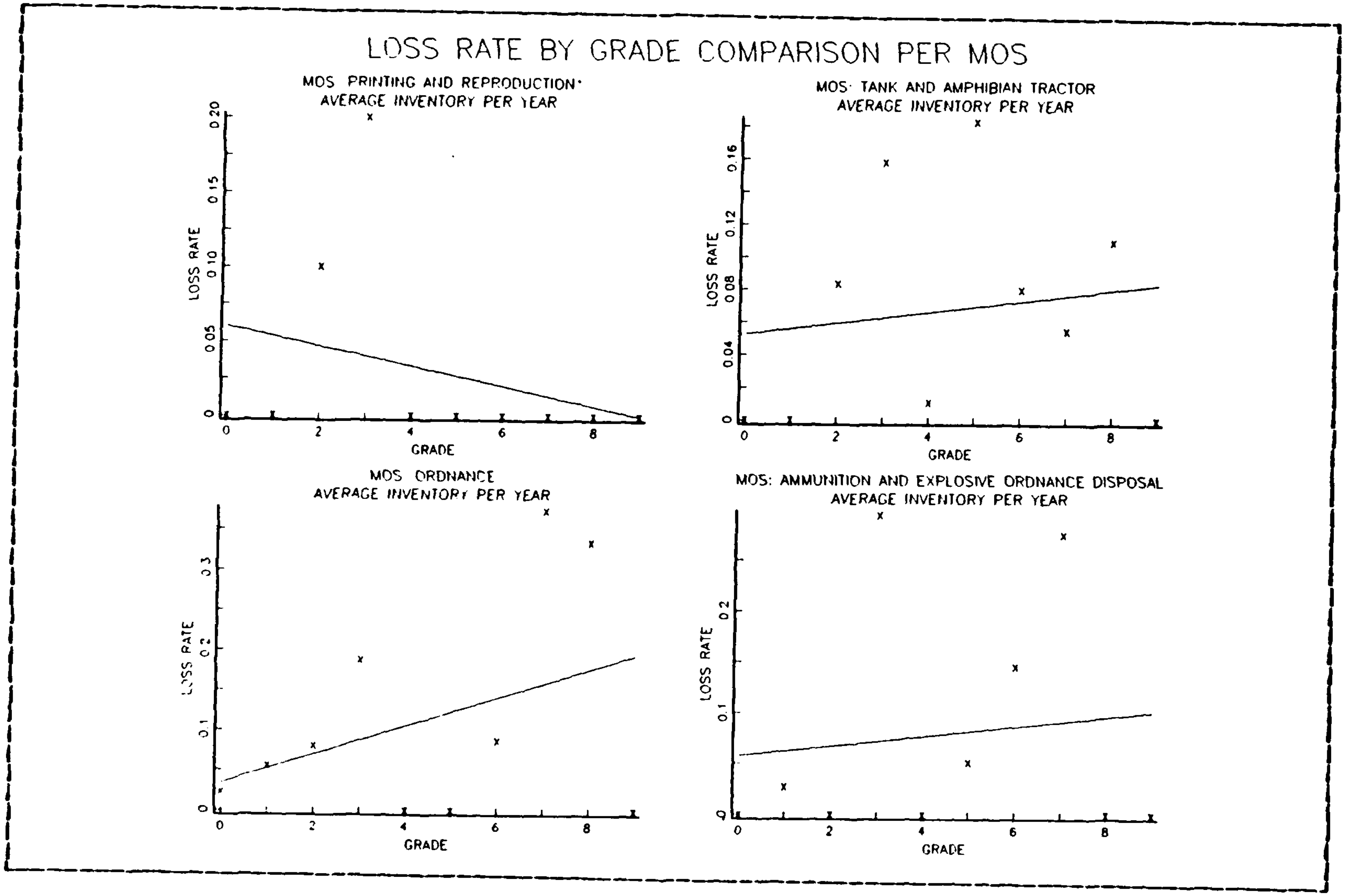


Figure A.13 Attrition Rates by Grade for MOS's 09 to 12.

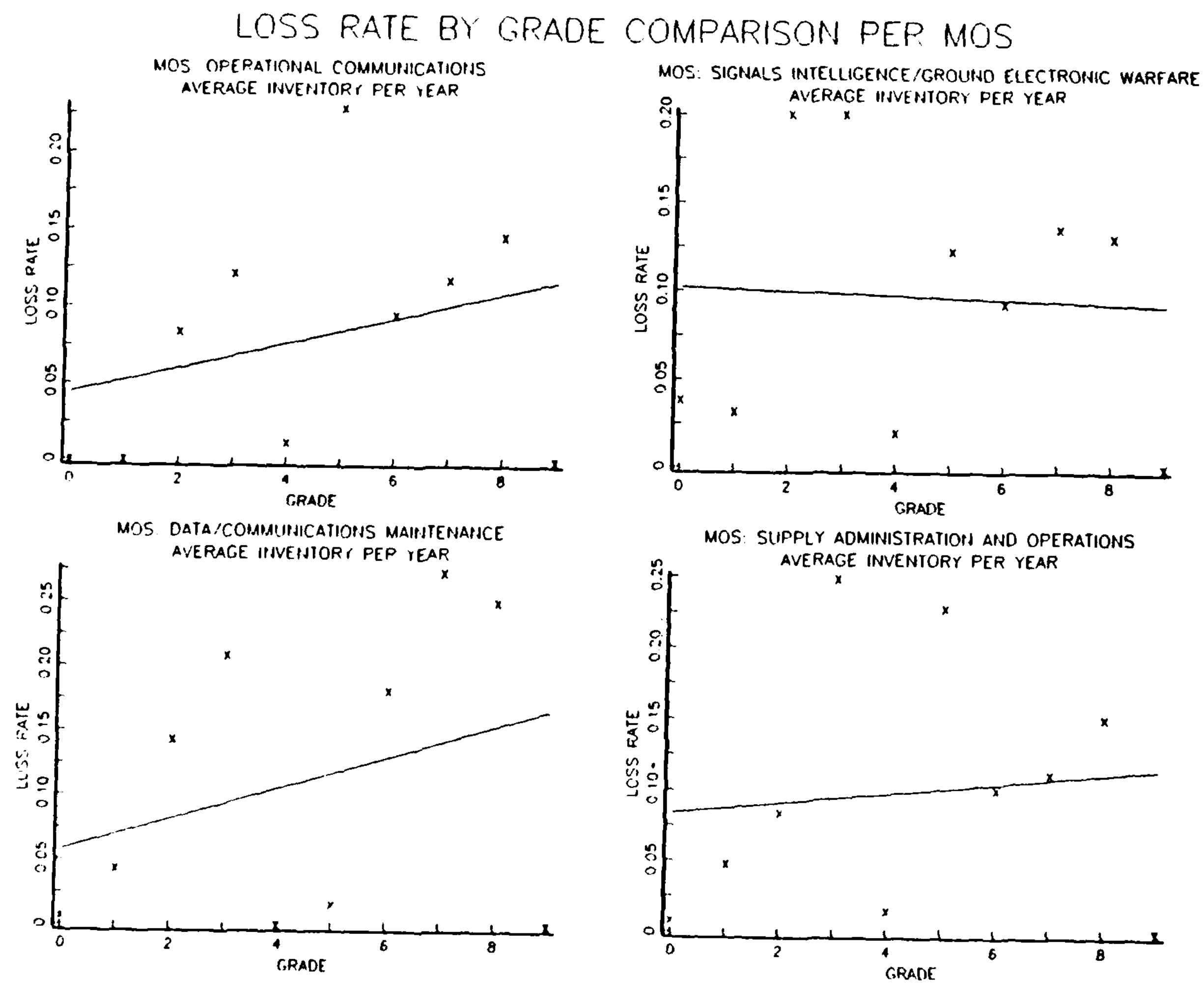


Figure A.14 Attrition Rates by Grade for MOS's 13 to 16.

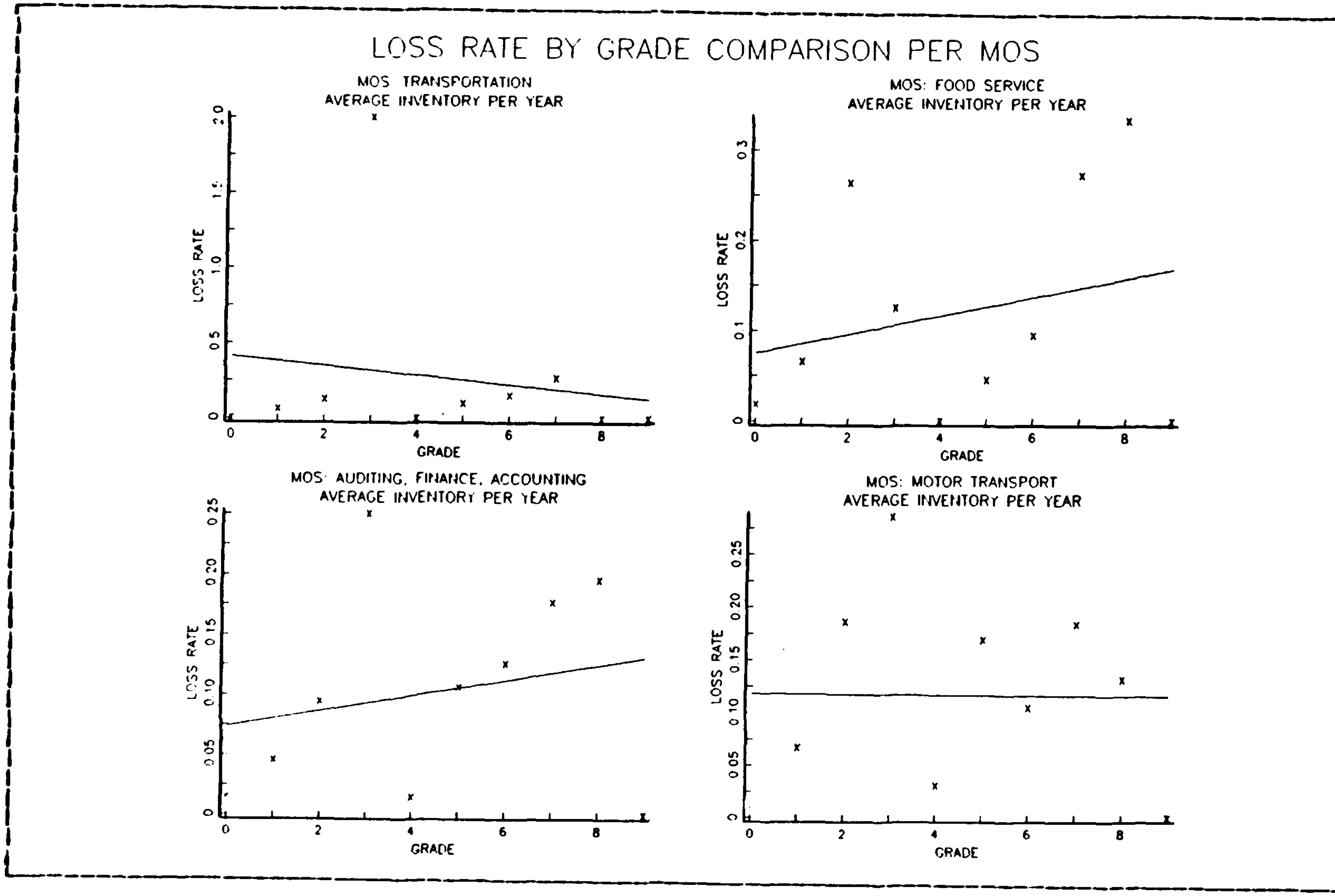


Figure A.15 Attrition Rates by Grade for MOS's 17 to 20.

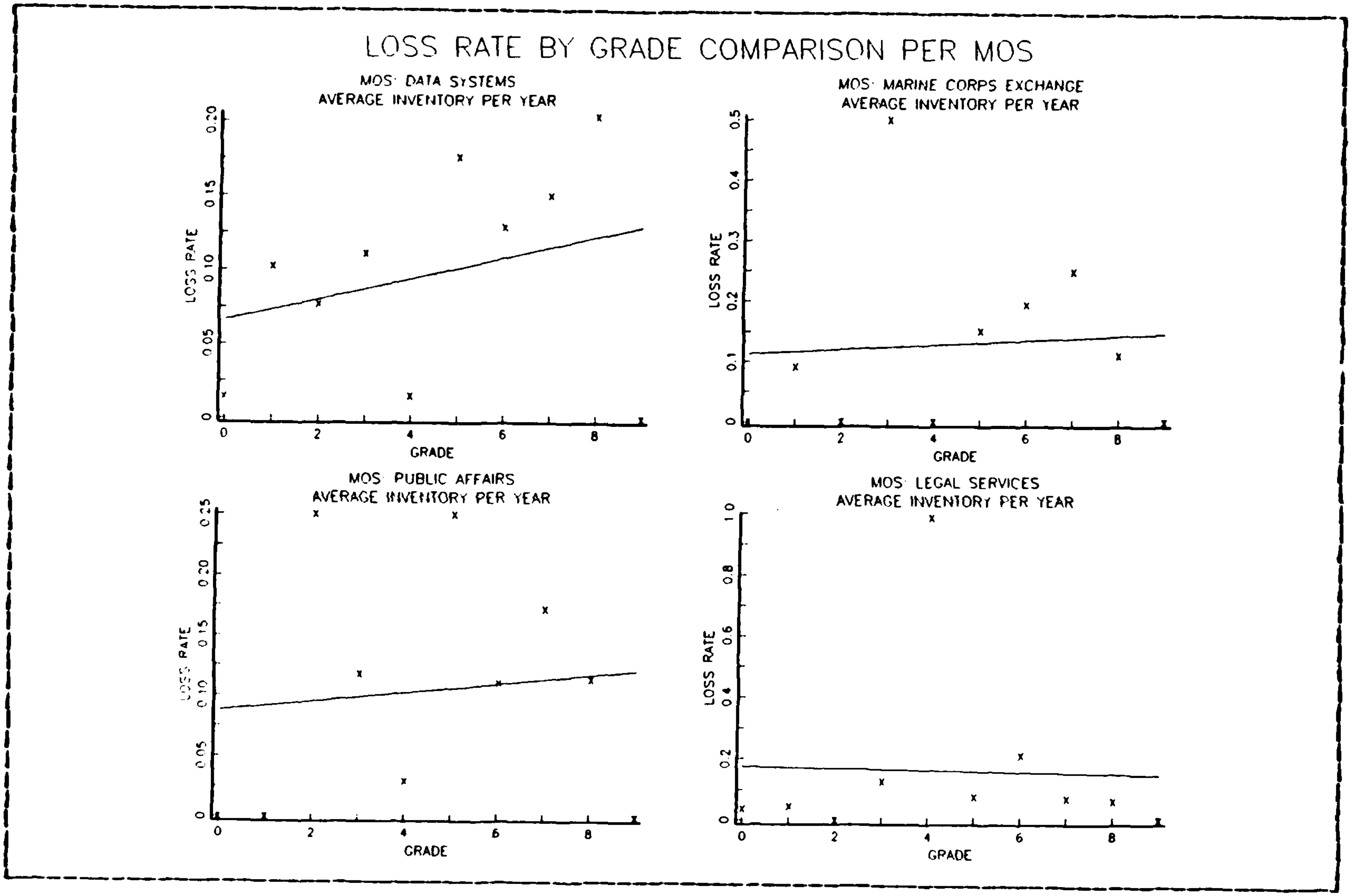


Figure A.16 Attrition Rates by Grade for MOS's 21 to 24.

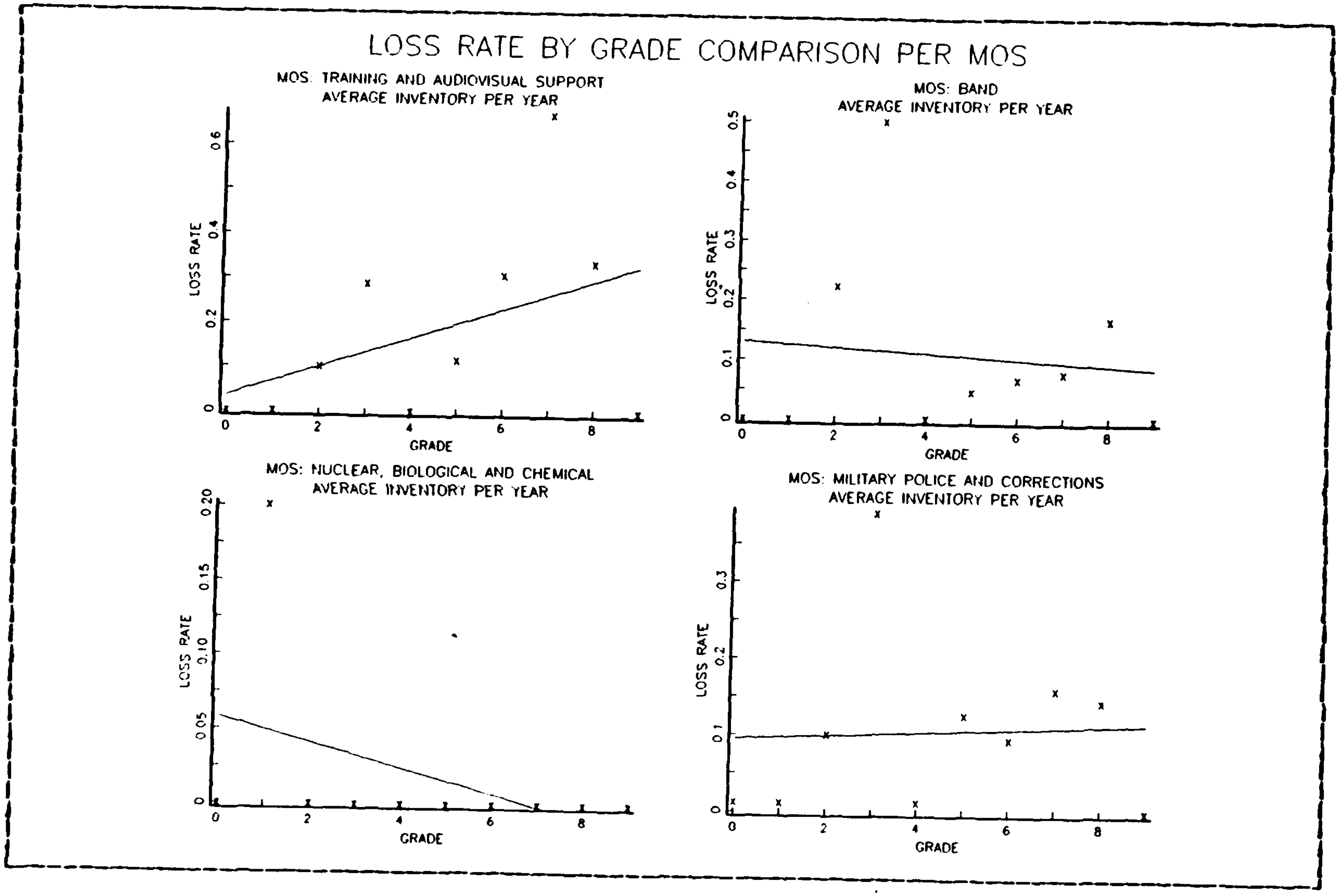


Figure A.17 Attrition Rates by Grade for MOS's 25 to 28.

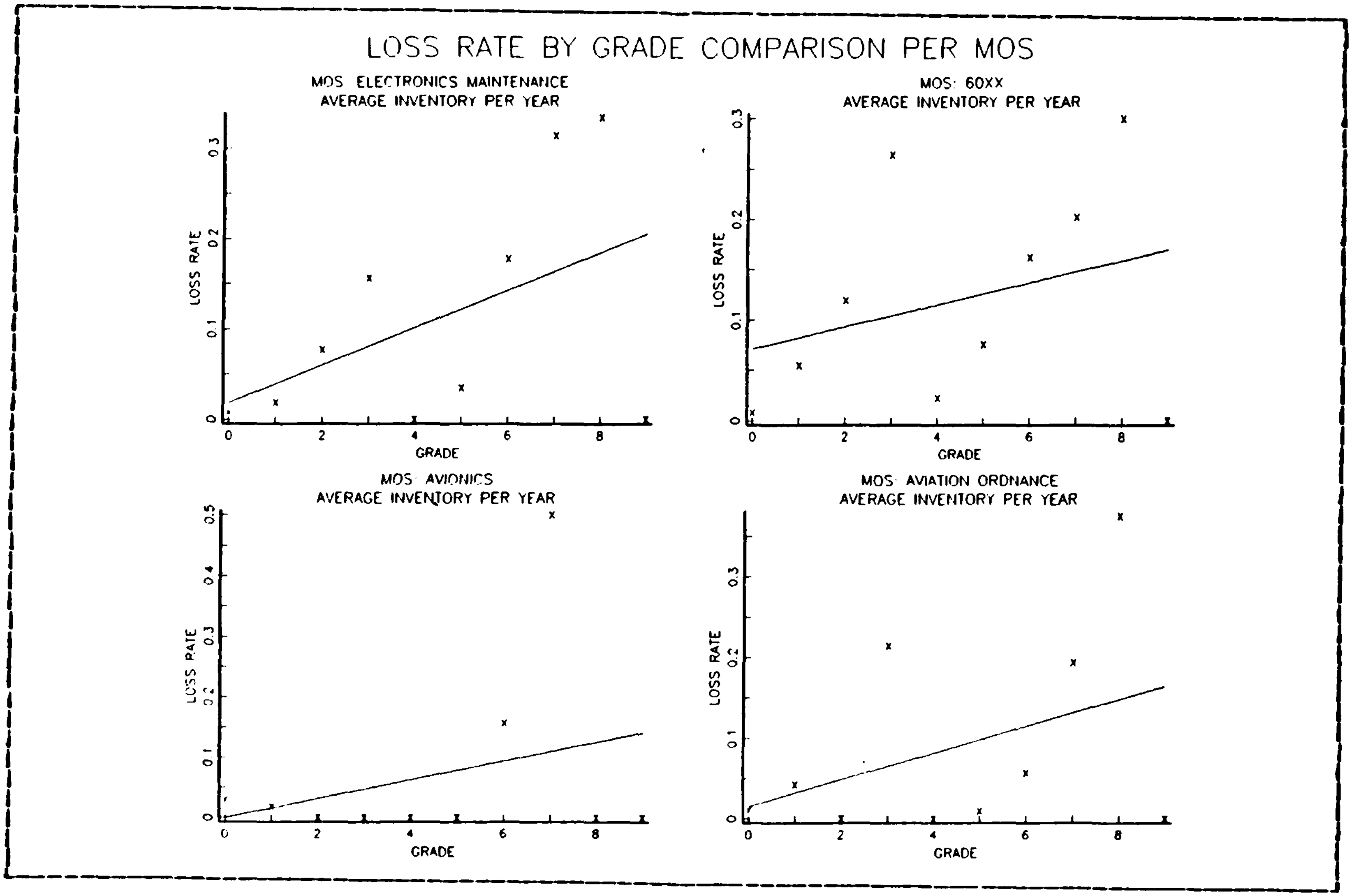


Figure A.18 Attrition Rates by Grade for MOS's 29 to 33.

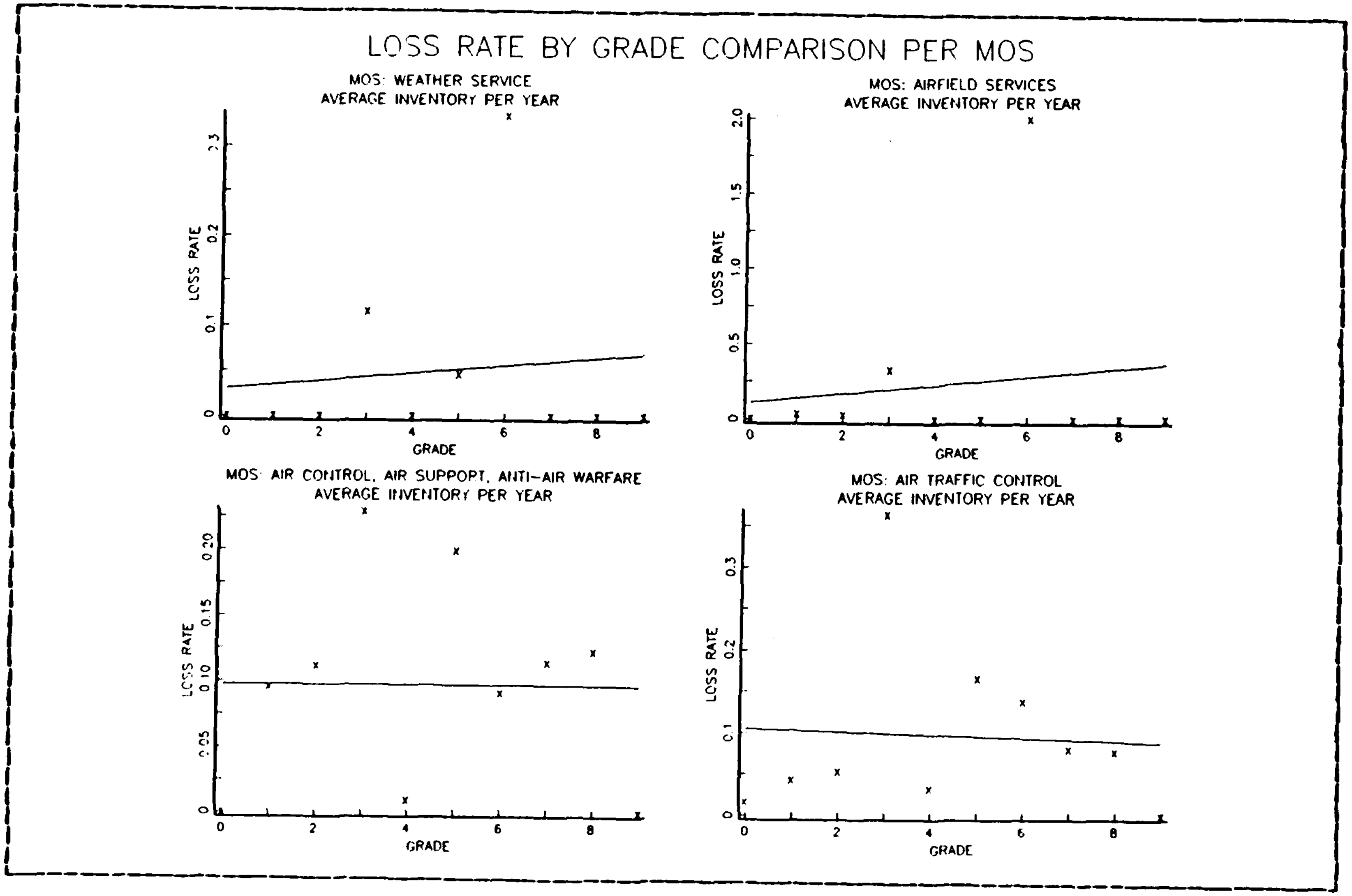


Figure A.19 Attrition Rates by Grade for MOS's 34 to 37.

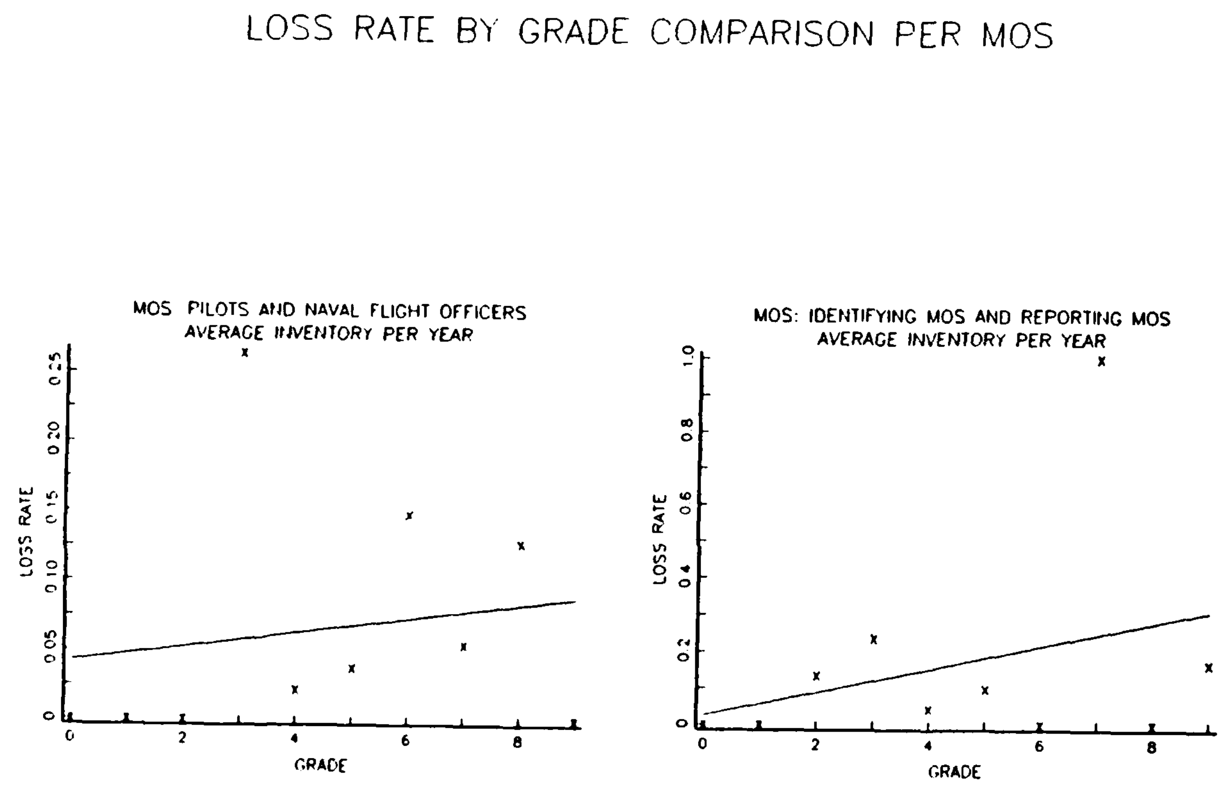


Figure A.20 Attrition Rates by Grade for MOS's 38 to 39.

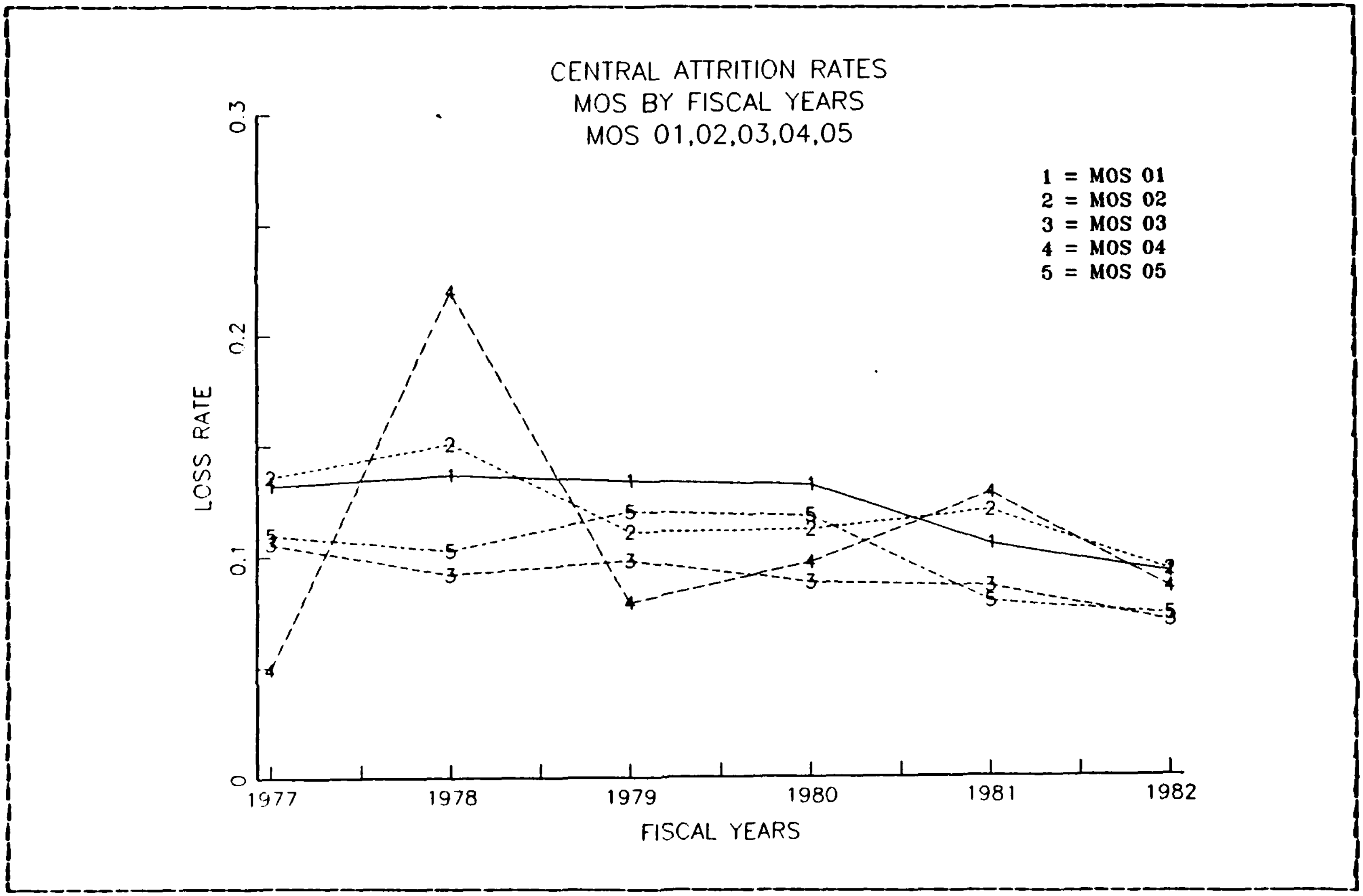


Figure A.21 Attrition Rates by Year for MOS's 01 to 05.

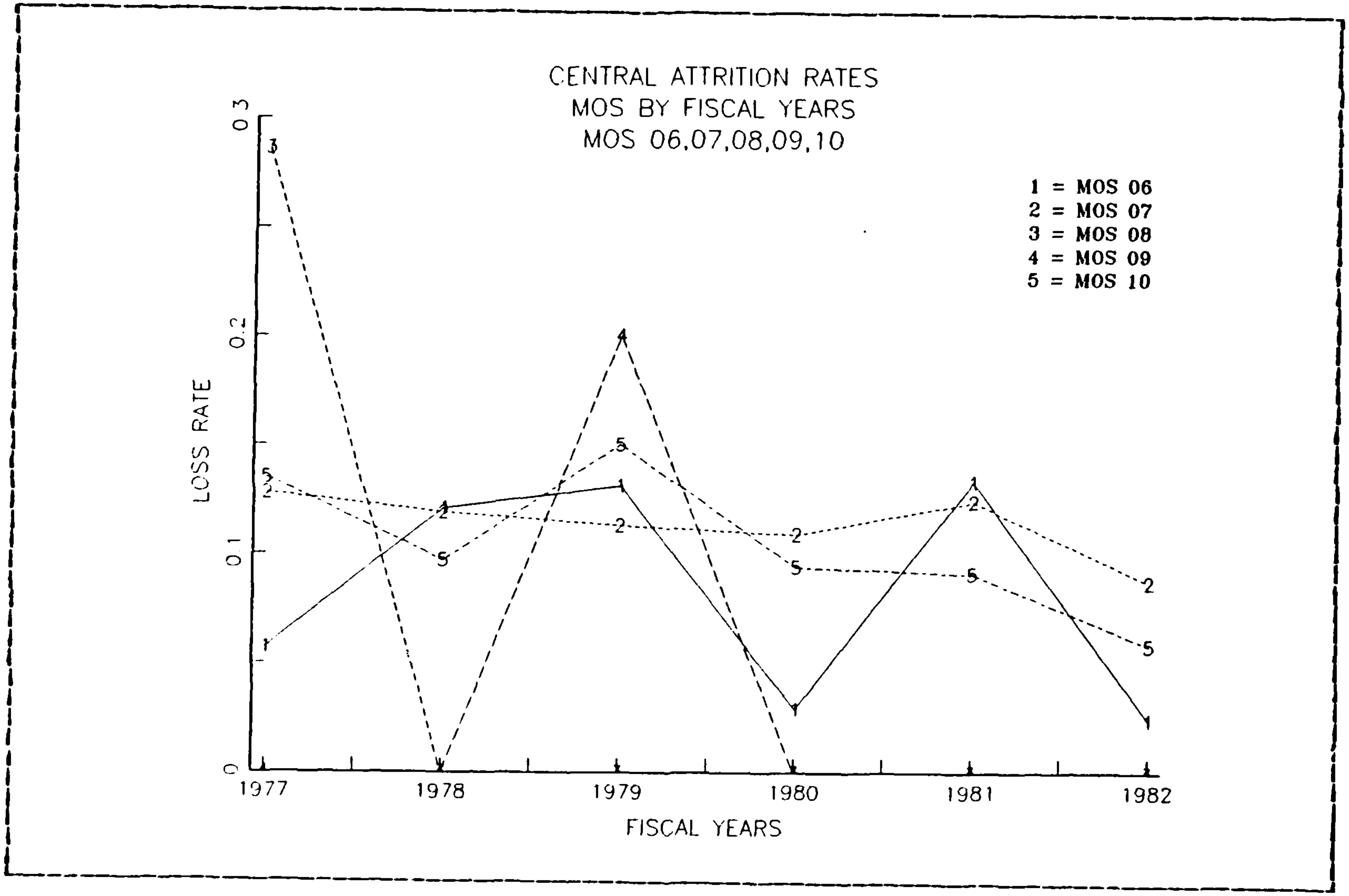
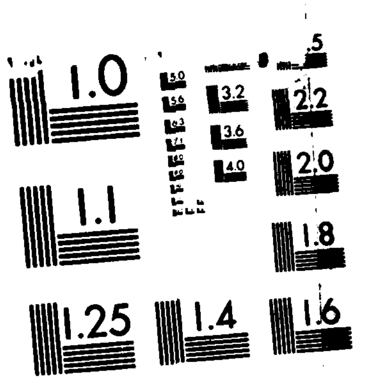


Figure A.22 Attrition Rates by Year for MOS's 06 to 10.

AD-A166 397 LOSS RATE ESTIMATION IN MARINE CORPS OFFICER MANPOWER MODELS(U) NAVAL POSTGRADUATE SCHOOL MONTEREY CA D D TUCKER SEP 85 2/2

UNCLASSIFIED F/G 5/9 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

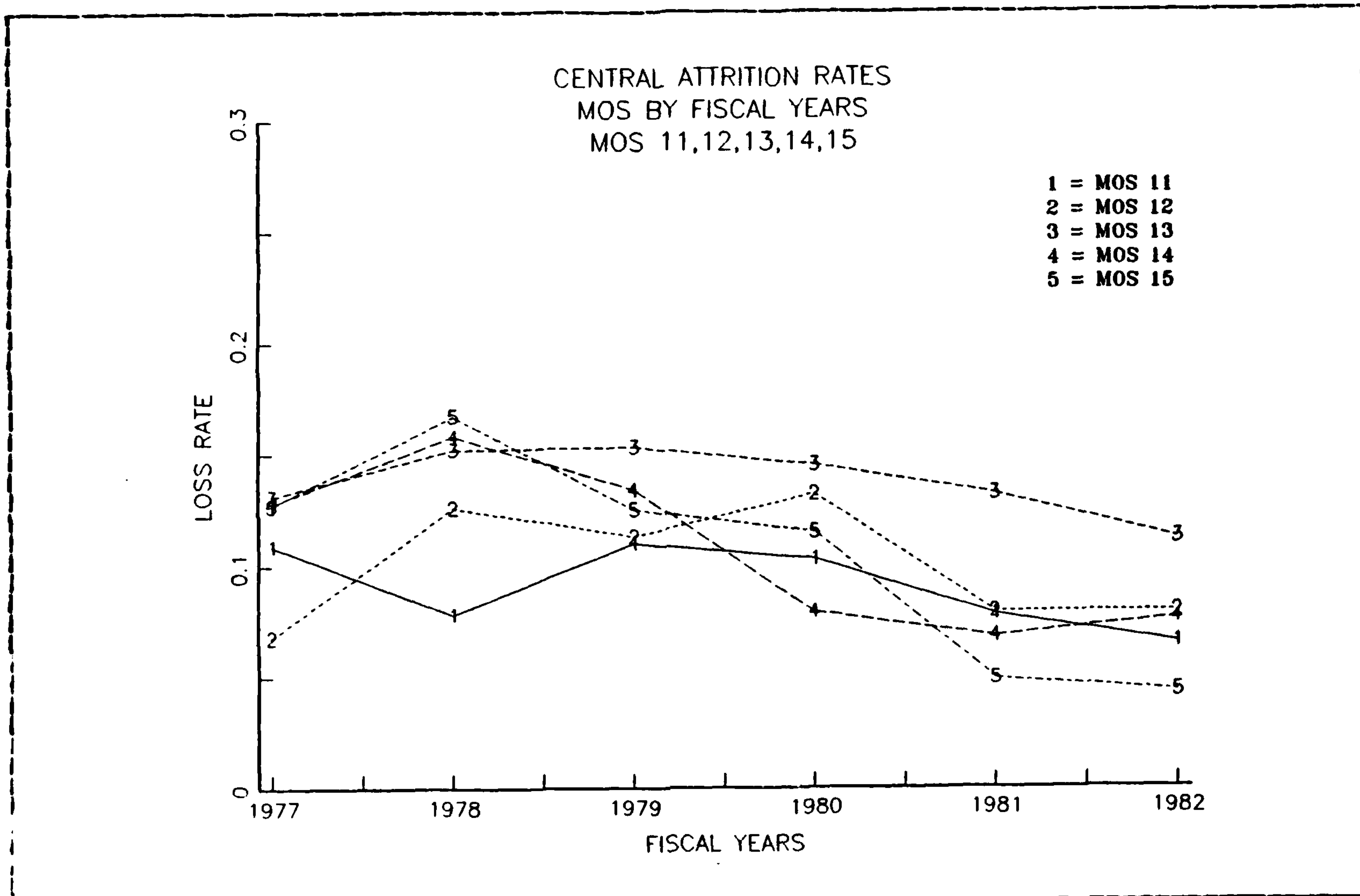


Figure A.23 Attrition Rates by Year for MOS's 11 to 15.

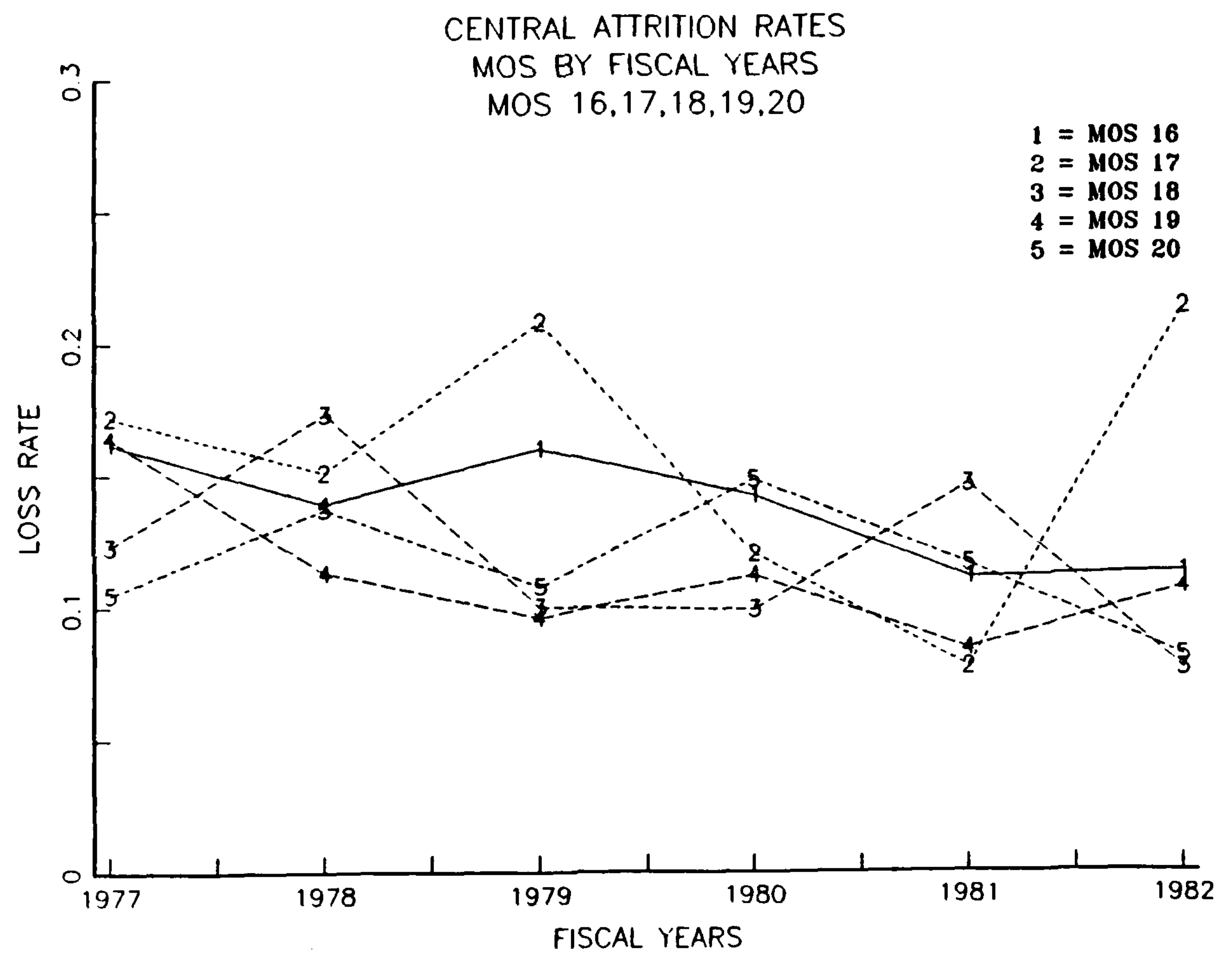


Figure A.24 Attrition Rates by Year for MOS's 16 to 20.

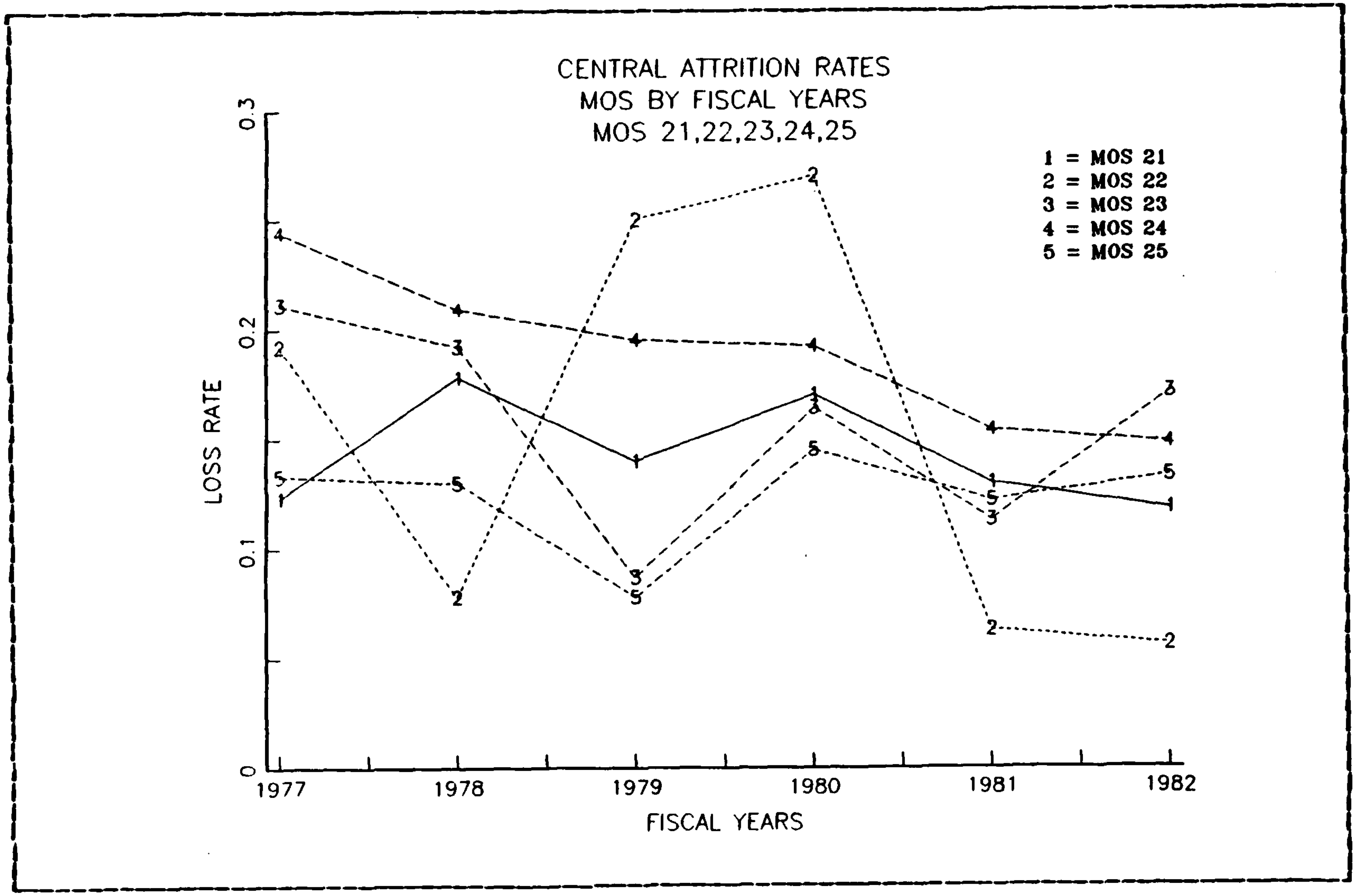


Figure A.25 Attrition Rates by Year for MOS's 21 to 25.

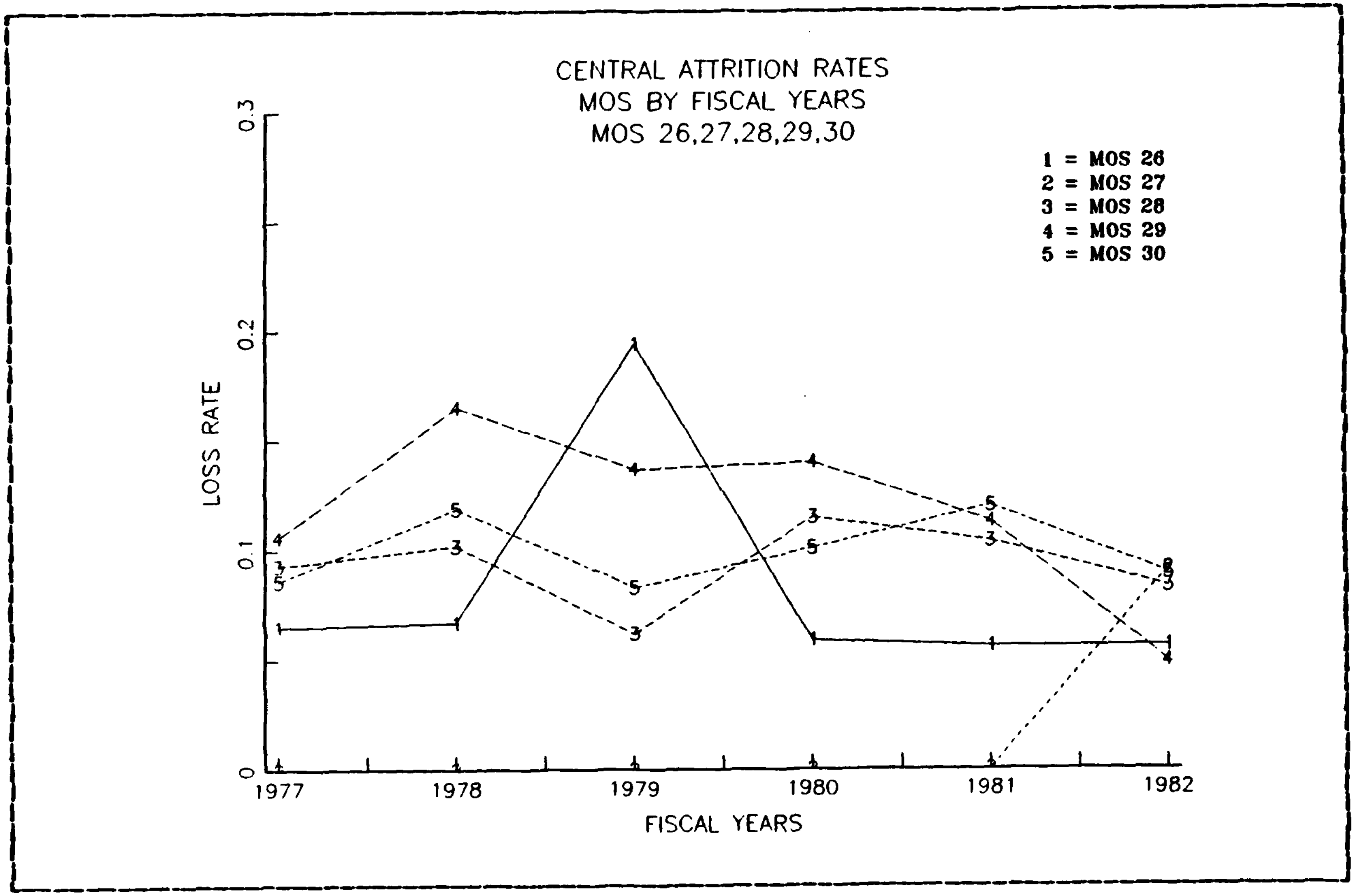


Figure A.26 Attrition Rates by Year for MOS's 26 to 30.

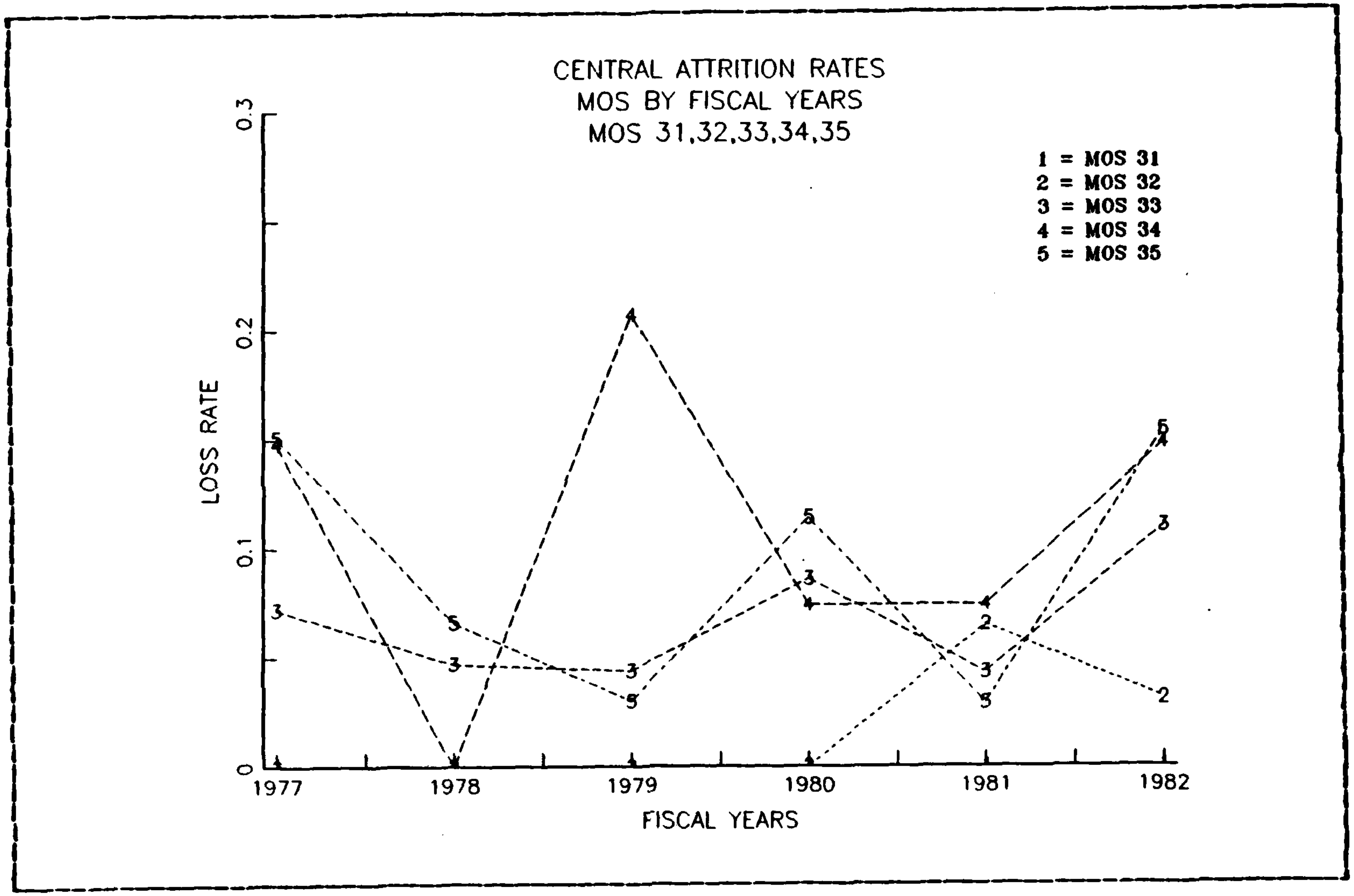


Figure A.27 Attrition Rates by Year for MOS's 31 to 35.

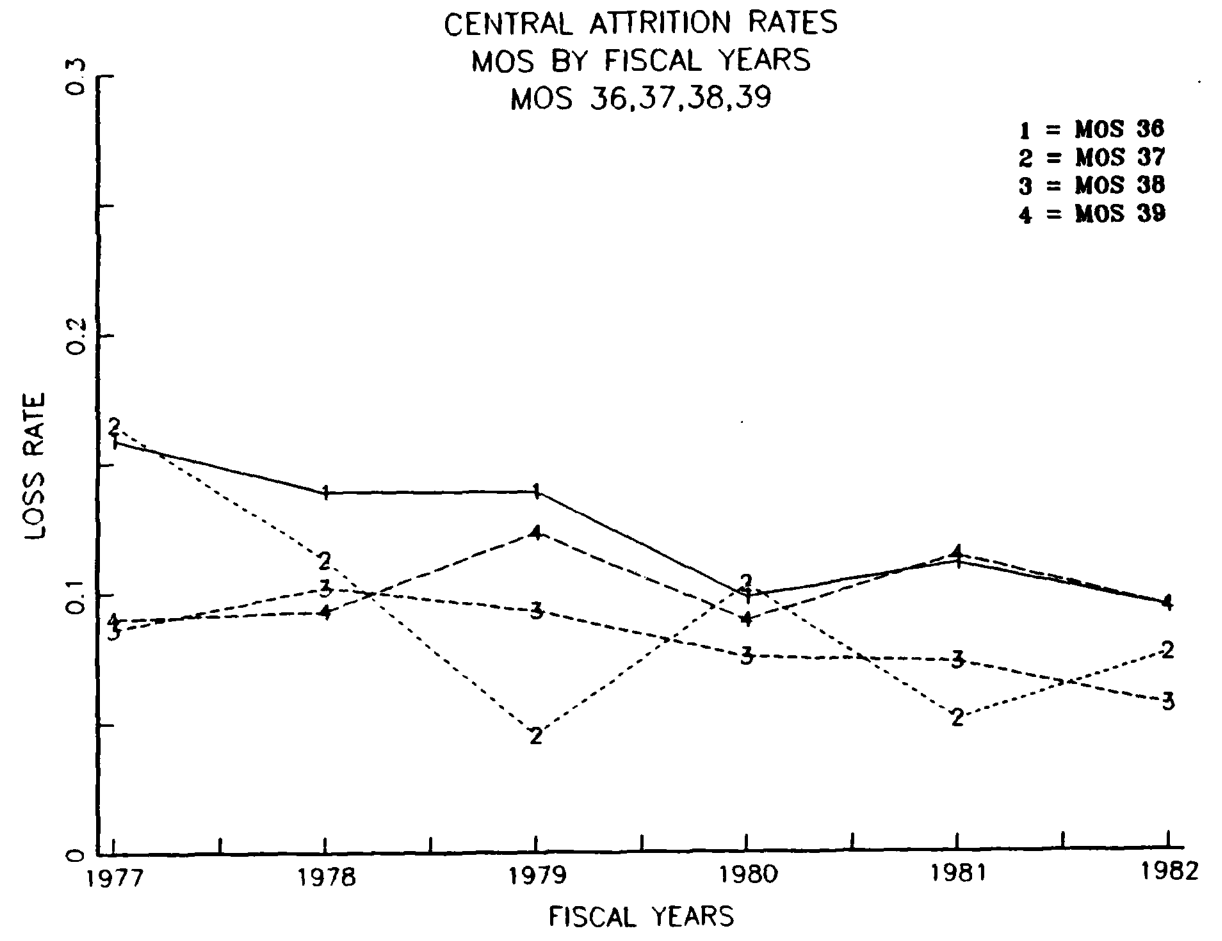


Figure A.28 Attrition Rates by Year for MOS's 36 to 39.

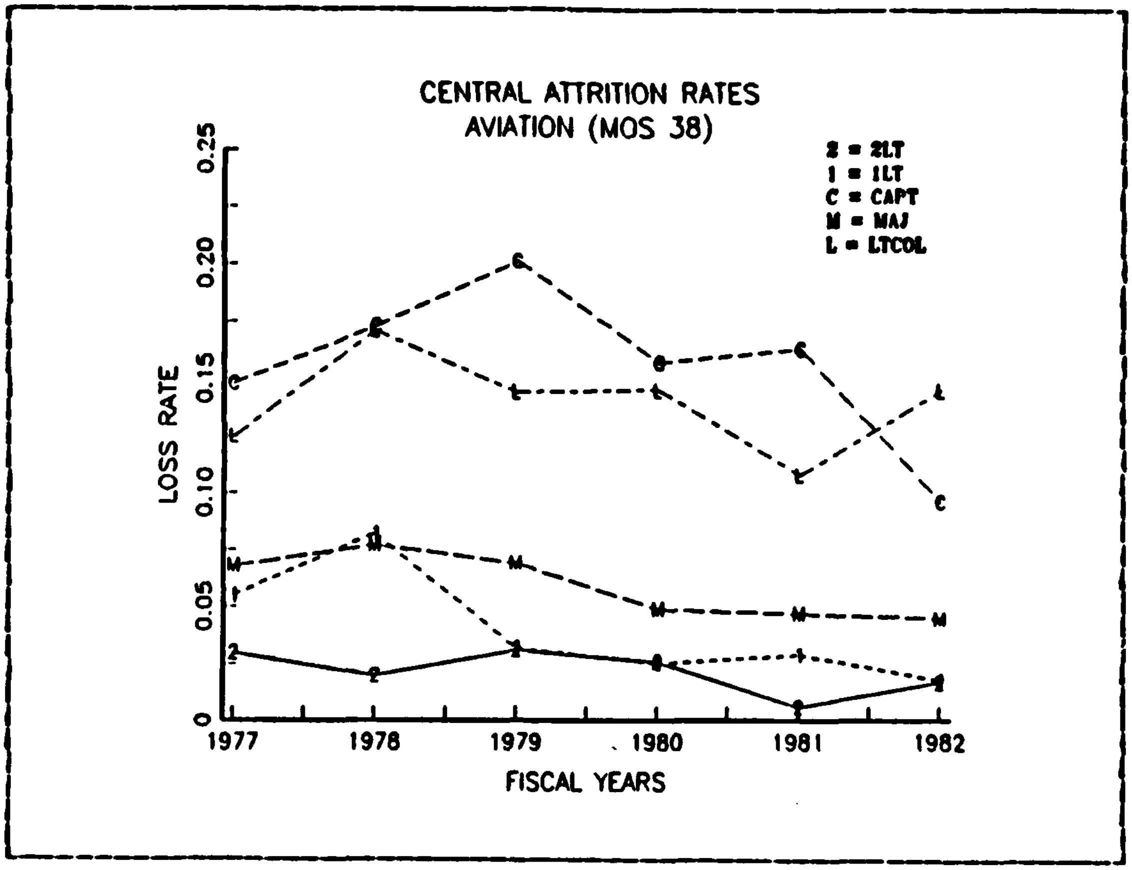


Figure A.29 Central Attrition Rates for the Aviation Group.

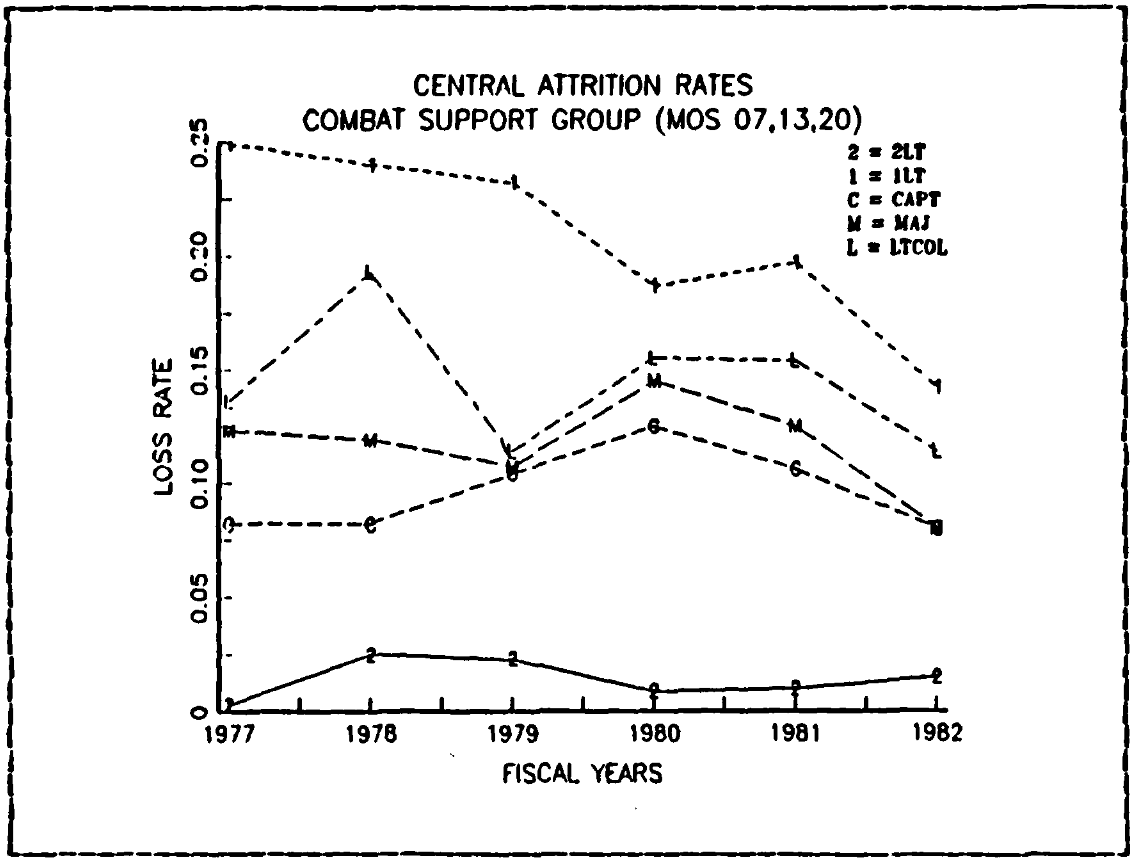


Figure A.30 Central Attrition Rates for the Combat Support Group.

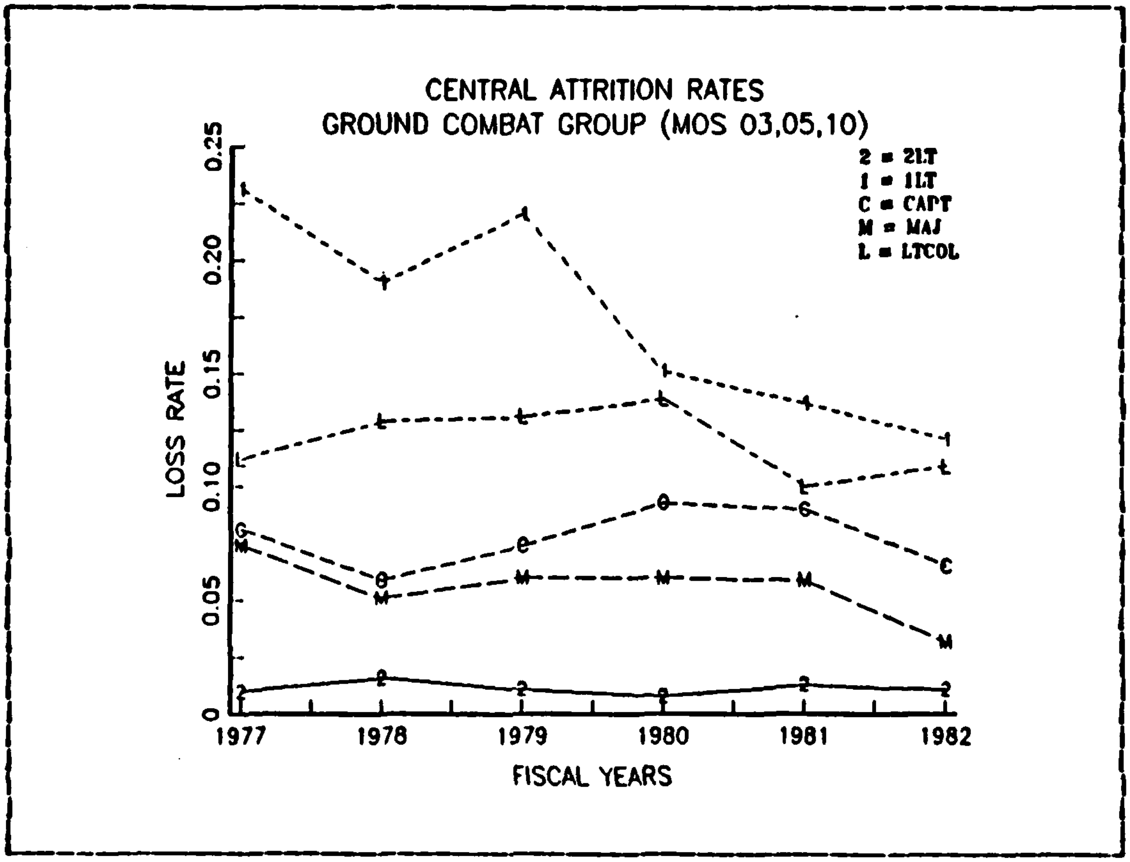


Figure A.31 Central Attrition Rates for the Ground Combat Group.

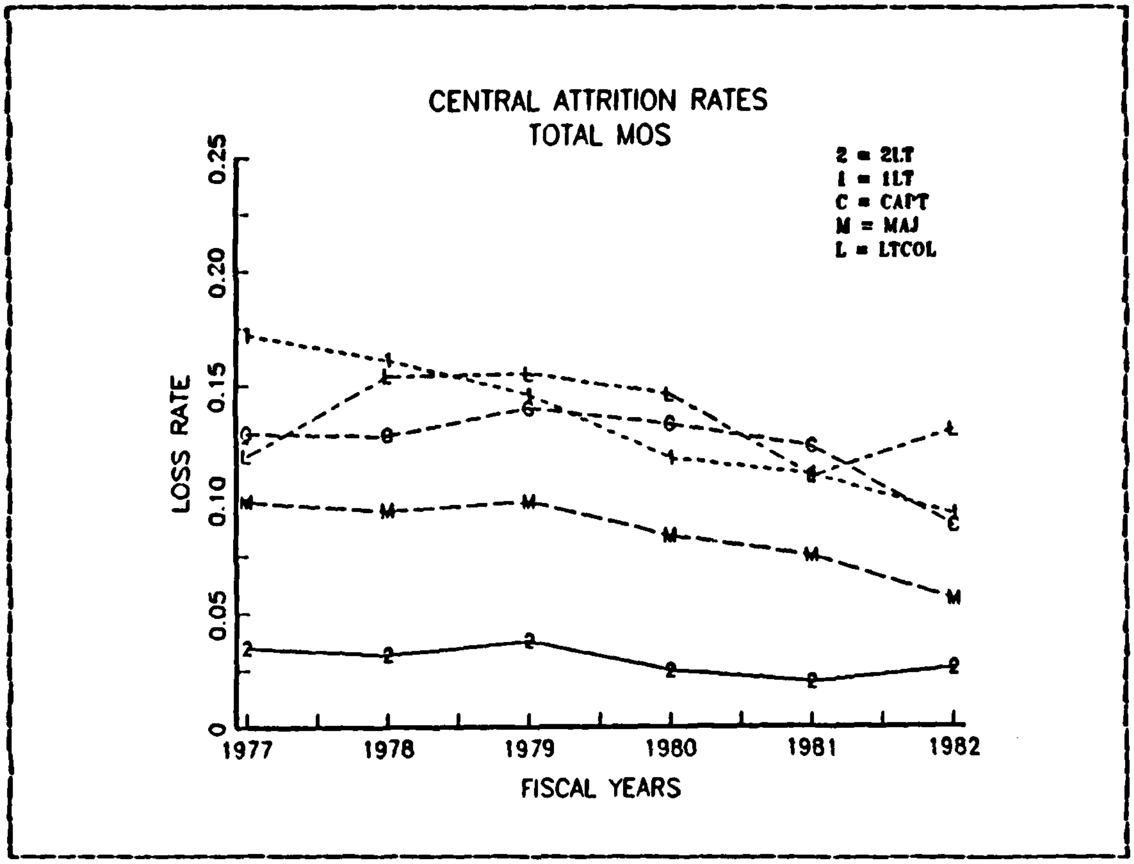


Figure A.32 Central Attrition Rates for the Total MOS's.

APPENDIX B

JAMES-STEIN ESTIMATOR ALGORITHM

This appendix contains the detailed algorithm for the James-Stein estimator of attrition rates for the USMC officer force structure. The information provided includes:

- Assumptions.
- Notation.
- Steps to generate the James-Stein attrition rates.

Algorithm: James-Stein Estimator of Attrition Rates. USMC Officers.

Segmentation: Hold each grade fixed and aggregate the MOS variable into four main cells. The four main cells are as follows:

1. Aviation
2. Ground Combat
3. Combat Support
4. Combat Service Support (all others)

Notation:

Let J = the number of MOS cells in the chosen aggregate.
Let I = the number of LOS cells in the chosen aggregate (usually 31).

Because the data is a combination of two types of data (central and transition) and because the inventory contains both structural and sampling zeroes, the following policies were adopted:

1. Let, $INV_{ij}(t)$ = inventory with LOS = i and MOS = j at the beginning of year t (t=1,...,T)

2. Let. $y_{ij}(t)$ = number of attritions in cell (i,j) at any time during year t.

3. Let, $n_{ij}(t)$ = maximum $\{y_{ij}(t)\ ,\ 0.5\ (INV_{ij}(t) + INV_{ij}(t+1))\}$

The incidence matrix D identifies the cells with non zero inventory.

Let, D_{ij} = 0 if $n_{ij}(t) = 0$ for all t=1,...,T
Let, D_{ij} = 1 otherwise.

The following steps are utilized to generate James-Stein attrition rates.

STEP 1: Use a variance stabalizing transform (Freeman - Tukey).

$$X_{ij}(t) = [n_{ij}(t) + \frac{1}{2}]^{\frac{1}{2}} \cdot (0.5)\ (\ Sin^{-1}[-1 + \frac{(2y_{ij}(t))}{(n_{ij}(t)+1)}]\ )$$

$$+\ (\ Sin^{-1}[\ -1 + \frac{2(y_{ij}(t)+1)}{(n_{ij}(t)+1)}]\ ).$$

STEP 2: Form the cell means and the grand mean.

$$\bar{X}_{ij} = \frac{1}{T}\sum_{t=1}^{T} X_{ij}(t) \qquad \text{For all (i,j) such that } D_{ij} = 1$$

$$\bar{X} = \frac{1}{K}\sum_{i=1}^{I}\sum_{j=1}^{J} \bar{X}_{ij} \cdot D_{ij}$$

$$K = \sum_{i=1}^{I}\sum_{j=1}^{J} D_{ij}$$

STEP 3: Form SSE, sum of squares error and SSB. sum of squares between by subtracting from SST, the total sum of squares.

$$SST = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T} (X_{ij}(t) - \bar{X})^2 \cdot D_{ij}$$

$$SSE = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=1}^{T}[X_{ij}(t) - \overline{X}_{ij}]^2 \cdot D_{ij}$$

$$SSB = SST - SSE$$

STEP 4: Compute the set of James-Stein estimators in the transformed scale.

$$C = 1 - \frac{T(K-3)}{K(T-1)+2}\,\frac{SSE}{SSB}$$

$$P = \begin{cases} \overline{X} + C(\overline{X}_{ij} - \overline{X}) & \text{if } D_{ij} = 1 \\ undefined & otherwise \end{cases}$$

STEP 5: Invert the transform to produce the attrition rates r_{ij}.

$$r_{ij} = \begin{cases} \frac{1}{2}[1 + \sin(\frac{P}{n_{ij}} + 0.5)] & \text{if } -\frac{\pi}{2} < \frac{P}{n_{ij}} + \frac{1}{2} < \frac{\pi}{2} \\ 0 & \text{if } -\frac{\pi}{2} \geqslant \frac{P}{n_{ij}} + \frac{1}{2} \\ 1 & \text{if } \frac{\pi}{2} \leqslant \frac{P}{n_{ij}} + \frac{1}{2} \end{cases}$$

$$n_{ij} = \frac{1}{T}\sum_{t=1}^{T} n_{ij}(t)$$

APPENDIX C

FORMULAS AND PROOFS

This appendix contains various formulas and proofs for the James-Stein process utilized throughout this research project. It is divided into eight sections as follows:

A. Simple Stochastic Model for Central Data.

B. Supporting Distributions.

C. Distribution "V"

D. Non Central Chi Square

E. Distribution of "W" and Expectation of 1/W

F. Loss and Risk Functions

G. Evaluation of E(U/W)

H. Inverse Sine Transform

I. Monthly Verses Yearly Rates

A. SIMPLE STOCHASTIC MODEL FOR CENTRAL DATA

This section contains the nomenclature and characteristics of a simple stochastic model for central data. It deals with the joint distribution of cell inventory and number of attritions.

The following nomenclature will be defined in this section as follows:

N = Inventory during period, which is distributed *Poisson* (λ) .

Y = Number of losses.

m = Central attrition rate = $\frac{Y}{N}$.

The conditional probability of losses given inventory is:

$$f(Y \mid N=n) = \binom{n}{y} \Theta^{y} (1-\Theta)^{n-y} \qquad y = 0,\ldots,n$$

The joint probability function is:

$$f(y,n) = f(y \mid n)\, g(n) = \binom{n}{y} \Theta^{y} (1-\Theta)^{n-y} e^{-\lambda} \frac{\lambda^{n}}{n!} \qquad 0 \leqslant y \leqslant n \leqslant \infty$$

Let

$$h(y) = P\{Y=y\} = \sum_{n=y}^{\infty} f(y,n) = \Theta^{y} \frac{e^{-\lambda}}{y!} \sum_{n=y}^{\infty} \frac{1}{(n-y)!} (1-\Theta)^{n-y} \lambda^{n}$$

$$= \frac{(\lambda\Theta)^{y} e^{-\lambda}}{y!} \sum_{n=y}^{\infty} \frac{[(1-\Theta)\lambda]^{n-y}}{(n-y)!}$$

Let $z = n - y$.

$$= \frac{(\lambda\Theta)^{y}}{y!} e^{-\lambda} \sum_{n=y}^{\infty} \frac{[(1-\Theta)\lambda]^{z}}{z!}$$

$$= \frac{(\lambda\Theta)^{y}}{y!} e^{-\lambda\Theta}$$

Therefore , Y is *Poisson* ($\lambda\Theta$) , which is the marginal distribution.

Now, the estimated value for the central attrition rate is:

$$E\,m = E\,\frac{Y}{N} = E\,E\left(\frac{Y}{N} \mid N\right) = E\,\frac{1}{N} E(Y \mid N) = E\,\frac{1}{N} N\Theta = \Theta$$

The conditional probability of inventory given losses is:

$$P(N=n \mid Y=y) = g(n \mid y) = \frac{f(y,n)}{h(y)}$$

$$= \frac{\binom{n}{y} \Theta^{y} (1-\Theta)^{n-y} e^{-\lambda} \frac{\lambda^{n}}{n!}}{e^{-\Theta\lambda} \frac{(\Theta\lambda)^{y}}{y!}} \qquad 0 \leqslant y \leqslant n \leqslant \infty$$

$$= \frac{1}{(n-y)!} e^{-\lambda-\Theta\lambda} \left[\frac{(1-\Theta)^{n-y}}{\lambda^{y}}\right] \lambda^{n}$$

$$= e^{-\lambda(1-\Theta)} \frac{\lambda(1-\Theta)^{n-y}}{(n-y)!} \qquad n = y, y-1, \ldots$$

This is a shifted Poisson, therefore

N - y is *Poisson* (λ (1 - Θ)) .

Characteristics of N - y are:

MEAN: $E[N \mid Y=y] = E[N - y + y \mid Y = y]$

$$= E[N - y \mid Y = y] + E[Y \mid Y = y]$$

$$= \lambda(1-\Theta) + y$$

VARIANCE: $Var[N \mid Y=y] = Var[N-y \mid Y=y] = \lambda(1-\Theta)$

COVARIANCE: $Cov(Y,N) = E[Y\ N] - E[Y]\ E[N]$

$$= E[E(Y\ N \mid N] - (\Theta\ \lambda)(\lambda)$$

$$= E[N\ E(Y \mid N)] - \Theta\ \lambda^2$$

$$= E[N(N\ \Theta)] - \Theta\ \lambda^2$$

$$= E[N^2\ \Theta] - \Theta\ \lambda^2$$

$$= \Theta\ (E(N^2) - \lambda^2)$$

$$= \Theta\ (Var(N))$$

$$= \Theta\ \lambda$$

CORRELATION: $Cor = \frac{Cov(Y,N)}{SD(Y)SD(N)}$

$$= \frac{\Theta\ \lambda}{\sqrt{\Theta\ \lambda}\sqrt{\lambda}}$$

$$= \sqrt{\Theta}$$

Therefore.

N = Inventory in a cell is distributed *Poisson* (λ) .

Y = Losses in a cell is distributed *Poisson* ($\Theta\ \lambda$)

Y N is distributed *Binomial* (N . Θ) .

Θ = Central attrition rate.

N Y is distributed Y + *Poisson* (λ (1 - Θ)) .

B. SUPPORTING DISTRIBUTIONS.

The following sections of this appendix contain a series of proofs to the following statements based on the assumption: Y is distributed Normal (Θ, I) . where I is the Identity matrix.

i) $U = \dfrac{\Theta' Y}{||\Theta||}$ is distributed Normal $(||\Theta||, 1)$.

ii) $Z = Y - \dfrac{\Theta' Y \Theta}{||\Theta||^2} = Y - U\dfrac{\Theta}{||\Theta||}$ is distributed Normal $(0, I - \dfrac{\Theta\Theta'}{||\Theta||^2})$.

See section C for the following statements:

iii) $V = ||Z||^2$ is distributed $\chi^2_{(K-1)}$.

iv) $W = U^2 + V = ||Y||^2$.

PROOFS.

i) U is independent of all components of Z. In other words all $Cov(U, Z) = 0$. Proof is as follows.

$$Cov(U, Z) = Cov(U, Y - \frac{U\Theta}{||\Theta||}) = Cov(U, Y) - \frac{\Theta}{||\Theta||} Cov(U, U)$$

Now, $Cov(U, Y) = Cov(\dfrac{\Theta' Y}{||\Theta||}, Y) = \dfrac{1}{||\Theta||}\sum_{j=1}^{K}\Theta_j\, Cov(Y_j, Y) = \dfrac{\Theta}{||\Theta||}$.

Therefore, $Cov(U, Z) = \dfrac{\Theta}{||\Theta||} - \dfrac{\Theta}{||\Theta||} = 0$. Q.E.D.

ii) The distribution for Z is Normal and the parameters are shown as follows.

$$Z_i = Y_i - \sum_{j=1}^{K}\frac{\Theta_j Y_j \Theta_i}{||\Theta||^2}$$

$$E(Z) = E(Y) - \frac{\Theta\Theta'}{||\Theta||^2}E(Y) = \Theta - \frac{\Theta\Theta'\Theta}{||\Theta||^2} = 0$$

$$E(ZZ') - (EZ)(EZ)' = E(ZZ') - 0 = E(ZZ') = Cov(Z, Z)$$

$$Cov(Z, Z) = Cov(Y - \frac{U\Theta}{||\Theta||}, Y - \frac{U\Theta}{||\Theta||})$$

$$= Cov(Y, Y) - Cov(\frac{U\Theta}{||\Theta||}, Y) - Cov(Y, \frac{U\Theta}{||\Theta||}) + Cov(\frac{U\Theta}{||\Theta||}, \frac{U\Theta}{||\Theta||})$$

Now $Cov(Y, Y) = I$. where I is the identity matrix since Y is distributed Normal (Θ, I) .

Also.

$$Cov(\frac{U\Theta}{||\Theta||}, \frac{U\Theta}{||\Theta||}) = E(\frac{U\Theta}{||\Theta||})(\frac{U\Theta}{||\Theta||})' - E(\frac{U\Theta}{||\Theta||})E(\frac{U\Theta}{||\Theta||})'$$

$$= E(\frac{U\Theta}{||\Theta||}(\frac{U\Theta}{||\Theta||})') - \frac{\Theta}{||\Theta||}E(U)\frac{\Theta}{||\Theta||}E(U)'$$

$$= E(U^2\frac{\Theta\Theta'}{||\Theta||^2}) - \Theta\Theta'$$

$$= E(U^2)\frac{\Theta\Theta'}{||\Theta||^2} - \Theta\Theta'$$

$$= (1 + ||\Theta||^2)\frac{\Theta\Theta'}{||\Theta||^2} - \Theta\Theta'$$

$$= \Theta\Theta'(\frac{1+||\Theta||^2}{||\Theta||^2} - 1)$$

$$= \Theta\Theta'(\frac{1}{||\Theta||^2} + 1 - 1)$$

$$= \frac{\Theta\Theta'}{||\Theta||^2}$$

Since $Cov(Y, \frac{U\Theta}{||\Theta||})$ and $Cov(\frac{U\Theta}{||\Theta||}, Y)$ are transposes of each other and covariance matrices are symmetric, the result is as follows using a previous result for $Cov(U, Y)$.

$$-2Cov(\frac{U\Theta}{||\Theta||}, Y) = -2\frac{\Theta}{||\Theta||}Cov(U, Y) = -2\frac{\Theta}{||\Theta||}(\frac{\Theta}{||\Theta||})'$$

Therefore,

$$Cov(Z, Z) = I - 2\frac{\Theta\Theta'}{||\Theta||^2} + \frac{\Theta\Theta'}{||\Theta||^2} = I - \frac{\Theta\Theta'}{||\Theta||^2}$$

and Z is distributed: Normal $(0, I - \frac{\Theta\Theta'}{||\Theta||^2})$.

C. DISTRIBUTION "V"

This section is an continuation of section B and shows $V = Z'Z = ||Z||^2$ is distributed $\chi^2_{(K-1)} = \sum_1^{K-1} Normal\,(0,1)^2$.

PROOFS.

iii) First, construct an orthonormal transformation P, where P is K by K dimensional and $P'P = I$, where I is the identity matrix.

Let p_K' be the last row of P and be defined by

$$p_K = \frac{\Theta}{||\Theta||}$$

$$||p_K||^2 = 1$$

Note that this can be done because an orthonormal tranformation can always be completed.

Now let $X = PZ$ and calculate

$$X_K = p_K'Z$$

$$= \sum_{j=0}^{K} \frac{\Theta_j Z_j}{||\Theta||}$$

$$= \frac{1}{||\Theta||} \sum_{j=0}^{K} \Theta_j \left(Y_j - \frac{U\Theta_j}{||\Theta||} \right)$$

$$= \frac{1}{||\Theta||} \left[\sum_{j=0}^{K} \Theta_j Y_j - \frac{U}{||\Theta||} \sum_{j=0}^{K} \Theta_j^2 \right]$$

$$= \frac{1}{||\Theta||} \left[U\,||\Theta|| - \frac{U\,||\Theta||^2}{||\Theta||} \right]$$

$$= 0$$

Therefore $X_K \equiv 0$ random variable, and the space is K-1 vice K dimensional.

Next calculate

$$Cov(X,X) = Cov(PZ,PZ) = E(PZ)(PZ)'$$

$$= E(PZZ'P') = P[E(ZZ')]P'$$

$$P(Cov(Z,Z))P' = P\left(I - \frac{\Theta\Theta'}{||\Theta||^2}\right)P'$$

$$= PP' - P\frac{\Theta\Theta'}{||\Theta||^2}P' = I - P\,p_K\,p_K'\,P'$$

Now,

$$P\,p_K\,p_K'\,P' = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}_{KxK}$$

Therefore,

$$Cov\,(X,X) = \begin{bmatrix} I_{K-1} & 0 \\ 0 & 0 \end{bmatrix}$$

Since $X = PZ$ implies $Z = P'X$, and

$$V = (Z'Z) = (P'X)'(P'X) = (X'PP'X) = X'X,$$

so V has same distribution as $||X||^2$.

Then,

$$V = \sum_{j=0}^{K} X_j^2 = \sum_{j=1}^{K-1} X_j^2$$

Therefore, V is distributed $\chi^2_{(K-1)}$.

To show

iv) $\quad W = U^2 + V = ||Y||^2$

$$U^2 + V = U^2 + ||Y - \frac{U\Theta}{||\Theta||}||^2$$

$$= U^2 + (Y - \frac{U\Theta}{||\Theta||})'(Y - \frac{U\Theta}{||\Theta||})$$

$$= U^2 + Y'Y - \frac{U\Theta'Y}{||\Theta||} - Y'\frac{U\Theta}{||\Theta||} + U^2\frac{\Theta'\Theta}{||\Theta||^2}$$

$$= U^2 + Y'Y - U^2 - U^2 + U^2$$

$$= Y'Y = ||Y||^2.$$

D. NON-CENTRAL CHI SQUARE

This section considers the Non-Central Chi Square with n degrees of freedom. Letting $Y_1, \ldots, Y_n$ be iid *Normal* (0,1), then $\chi^2_{(n)} = \sum_{i=1}^{n} Y_i^{\,2}$ with density:

$$f_n(u) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1} e^{-\frac{u}{2}} \qquad 0 < n < \infty$$

But if the means are not all zero then the result is the Non-Central Chi Square random variable.

i) With n=1 and letting Y be distributed *Normal* (Θ,1) and $V = Y^2$, then the density of V is as follows.

$$y^2 = v \qquad y = \pm\sqrt{v} \qquad \left|\frac{dy}{dv}\right| = \frac{d\,|\,v^{\frac{1}{2}}\,|}{dv} = \frac{1}{2}v^{-\frac{1}{2}} = \frac{1}{2\sqrt{v}}$$

$$f_V(v) = [f_Y(\sqrt{v}\,) + f_Y(-\sqrt{v}\,)]\,\frac{1}{2\sqrt{v}} \qquad 0 < v < \infty$$

$$= \frac{1}{2\sqrt{v}}\frac{1}{\sqrt{2\pi}}\left(e^{-\frac{1}{2}[\sqrt{v}-\Theta]^2} + e^{-\frac{1}{2}[-\sqrt{v}-\Theta]^2}\right)$$

$$= \frac{1}{2\sqrt{v}}\frac{1}{\sqrt{2\pi}}\left(e^{-\frac{1}{2}[v-2\Theta\sqrt{v}+\Theta^2]} + e^{-\frac{1}{2}[v+2\Theta\sqrt{v}-\Theta^2]}\right)$$

$$= \frac{1}{2\sqrt{v}}\frac{1}{\sqrt{2\pi}} e^{-\frac{v}{2}} e^{-\frac{\Theta^2}{2}}\left(e^{\Theta\sqrt{v}} + e^{-\Theta\sqrt{v}}\right)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\Theta^2}{2}} \frac{1}{\sqrt{v}} e^{-\frac{v}{2}} Cosh\,(\Theta\sqrt{v}\,)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\Theta^2}{2}} e^{-\frac{v}{2}} \frac{1}{\sqrt{v}} \sum_{n=0}^{\infty} \frac{(\Theta\sqrt{v}\,)^{2n}}{(2n)!}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{\Theta^2}{2}} e^{-\frac{v}{2}} \frac{1}{\sqrt{v}} \sum_{n=0}^{\infty} (\frac{\Theta^2}{2})^n 2^n v^n \frac{n!}{n!}\frac{1}{(2n)!}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{v}{2}} \frac{1}{\sqrt{v}} \sum_{n=0}^{\infty} \frac{e^{-\frac{\Theta^2}{2}}}{n!} (\frac{\Theta^2}{2})^n \frac{2^n n!}{(2n)!} v^n$$

$$= \sum_{n=0}^{\infty} P_n(\lambda) v^{n-\frac{1}{2}} e^{-\frac{v}{2}} \frac{1}{\sqrt{2\pi}} \frac{2^n n!}{(2n)!}$$

Where

$$\lambda = \frac{\Theta^2}{2}, \qquad v^{n-\frac{1}{2}} e^{-\frac{v}{2}} = v^{\frac{2n+1}{2}-1} e^{-\frac{v}{2}}, \qquad P_n(\lambda) = e^{-\lambda}\frac{\lambda^n}{n!}$$

and $\frac{1}{\sqrt{2\pi}} \frac{2^n n!}{(2n)!}$ is the constant for the central $\chi^2_{\left(\frac{2n+1}{2}\right)}$ density function.

ii) The case of general "n" can be managed by transforming all of the non-centrality into the first variable. As before, $V = \sum_{i=1}^{n} Y_i^2$.

Let P be an orthonormal transformation $(P'P = I)$ and having the first column p_1:

$$p_1 = \frac{\Theta}{||\Theta||}$$

Note that $p_1'p_1 = \frac{\Theta'\Theta}{||\Theta||^2} = 1$ and that an orthonormal transformation can always be completed.

Then let $T = P'Y$

and hence $V = Y'Y = T'P'PT = T'T = ||T||^2$

and $T_1, \ldots, T_n$ are independent (i.e., since T is a normal random vector, then one needs only compute the covariance matrix).

$$Cov(T,T') = E(TT') - E(T)E(T')$$

Now for i=1

$$E(T_1) = E(p_i'Y) = \frac{\Theta'}{||\Theta||}E(Y) = \frac{\Theta'\Theta}{||\Theta||} = ||\Theta||$$

For $i > 1$

$$E(T_i) = E(p_i'Y) = p_i'E(Y) = p_i'\Theta = ||\Theta||\, p_i'p_1 = 0$$

Hence $E(T') = (||\Theta||, 0, \ldots, 0)$ and

$$E(TT') = E(P'YY'P) = P'E(YY')P$$

$$= P'[Cov(Y,Y') + E(Y)E(Y')]P$$

$$= P'[I + \Theta\Theta']P$$

$$= P'P + P'\Theta\Theta'P$$

Using $P'P = I$ and

$$\Theta'P = (||\Theta||, 0, \ldots, 0) = E(T')$$

Hence,

$$Cov(T,T') = I + E(T)E(T') - E(T)E(T') = I$$

and independence follows.

Since $E(T_i) = 0$ for $i > 1$, it follows that the Non-Central $\chi^2_{(n)}(\lambda)$ where $\lambda = \frac{1}{2}||\Theta||^2$ is the convolution of $\chi^2_{(1)}(\lambda)$ and a Central $\chi^2_{(n-1)}$. This is treated in the next section.

E. DISTRIBUTION of W and EXPECTATION of W^{-1}

This section illustrates the characteristics of the distribution "W", which is the convolution of $f_{U^2}(s)$ (non-central chi square) and the density of $\chi^2_{(K-1)} = f_{K-1}(u)$. It is noted that W is a non-central chi square random variable with K degrees of freedom and non-centrality parameter λ.

$$f_W(w) = \int_0^\infty f_{U^2}(w-v)\, f_{K-1}(v)\, dv$$

$$= \int_0^\infty \sum_{n=0}^\infty P_n(\lambda)\, f_{2n-1}(w-v)\, f_{K-1}(v)\, dv$$

$$= \sum_{n=0}^\infty P_n(\lambda) \int_0^\infty f_{2n-1}(w-v)\, f_{K-1}(v)\, dv$$

$$= \sum_{n=0}^\infty P_n(\lambda)\, f_{2n+K}(w)$$

Desire to derive

$$E\left[\frac{1}{W}\right] = \int_0^\infty \frac{1}{w} \sum_{n=0}^\infty P_n(\lambda)\, f_{2n+K}(w)\, dw$$

$$= \sum_{n=0}^\infty P_n(\lambda) \int_0^\infty \frac{1}{w}\, f_{2n+K}(w)\, dw$$

Let $r = 2n - k$

Now $f_r(w) = \dfrac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})}\, w^{\frac{r}{2}-1} e^{-\frac{u}{2}}$.

$$E\left[\frac{1}{W}\right] = \sum_{n=0}^\infty P_n(\lambda) \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} \int_0^\infty \frac{1}{w}\, w^{\frac{r}{2}-1} e^{-\frac{u}{2}}\, dw$$

$$= \sum_{n=0}^\infty P_n(\lambda) \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} \int_0^\infty w^{\frac{r}{2}-2} e^{-\frac{u}{2}}\, dw$$

Let $u = \dfrac{w}{2}$, then $du = \dfrac{1}{2}\, dw$

$$E\left[\frac{1}{W}\right] = \sum_{n=0}^\infty P_n(\lambda) \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} \int_0^\infty (2u)^{\frac{r}{2}-2} e^{-u}\, 2\, du$$

$$= \sum_{n=0}^{\infty} P_n(\lambda) \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} 2^{\frac{r}{2}-1} \int_0^{\infty} u^{\frac{r-2}{2}-1} e^{-u}\, du$$

$$= \sum_{n=0}^{\infty} P_n(\lambda) \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} 2^{\frac{r}{2}-1}\Gamma(\frac{r-2}{2})$$

$$= \sum_{n=0}^{\infty} P_n(\lambda)\ \frac{1}{2}\ \frac{\Gamma(\frac{r}{2}-1)}{(\frac{r}{2}-1)\Gamma(\frac{r}{2}-1)}$$

$$= \sum_{n=0}^{\infty} P_n(\lambda) \frac{1}{(r-2)}$$

$$= \sum_{n=0}^{\infty} P_n(\lambda) \frac{1}{(2n+K-2)}$$

$$= E\ \frac{1}{2J+K-2}$$

Where J is distributed Poisson (λ).

Now $\frac{1}{2J+K-2} \leq 1$ for $K \geq 3$.

Therefore,

$$E\ \frac{1}{W} = E\ \frac{1}{2J+K-2} \leq 1 \qquad \lambda \geq 0$$

F. LOSS AND RISK FUNCTIONS

This section evaluates the Loss and Risk functions. Considering the column vector Y, which is distributed Normal (Θ, I), these functions are as follows.

$$LOSS = L(a,\Theta) = (a-\Theta)(a-\Theta)' = ||a-\Theta||^2$$

Risk is expected loss, and for the maximum likelihood estimator, a = Y and

$$RISK = E[L(Y,\Theta)] = E(Y-\Theta)'(Y-\Theta) = E(||Y-\Theta||^2) = K$$

Consider another function of Y. Use $||Y||^2 = Y'Y = \sum_{j=0}^{K} Y_i^2$ which is distributed Non Central $\chi^2_{(K)}$. See section D for the properties of the Non Central Chi Square Distribution. The James-Stein estimator has structure:

$$a(Y) = (1-\frac{b}{||Y||^2})Y$$

It's Risk function is derived as follows:

$$RISK = E[||(1-\frac{b}{||Y||^2})Y-\Theta||^2]$$

$$= E[||Y-\Theta-\frac{bY}{||Y||^2}||^2]$$

$$= E[(Y-\Theta-\frac{bY}{||Y||^2})'(Y-\Theta-\frac{bY}{||Y||^2})]$$

$$= E[(Y-\Theta)-\frac{bY}{||Y||^2}]'[(Y-\Theta)-\frac{bY}{||Y||^2}]$$

$$= E[(Y-\Theta)'(Y-\Theta)-\frac{b}{||Y||^2}Y'(Y-\Theta)-\frac{b}{||Y||^2}(Y-\Theta)'Y-\frac{b^2}{||Y||^4}Y'Y]$$

$$= E[||Y-\Theta||^2-2\frac{b}{||Y||^2}Y'(Y-\Theta)-\frac{b^2}{||Y||^4}||Y||^2]$$

$$= E[||Y-\Theta||^2]-2E[\frac{b}{||Y||^2}Y'(Y-\Theta)]-b^2E[\frac{1}{||Y||^2}]$$

$$= K-2E[\frac{b}{||Y||^2}Y'(Y-\Theta)]+b^2E[\frac{1}{\chi^2_{(K)}(\lambda)}]$$

Let

$$EVAL = E(\frac{(Y-\Theta)'Y}{||Y||^2})$$

$$= 1-||\Theta||E(\frac{U}{W})$$

where U and W are defined in section B of this appendix.

See section G in this appendix for an evaluation of $E(\frac{U}{W})$. Using that result now leads to the following:

$$EVAL = 1 - \|\Theta\|^2 \sum_{j=0}^{\infty} P_j(\lambda) \frac{1}{K-2j}$$

Where $\lambda = \frac{\|\Theta\|^2}{2}$, and $P_j(\lambda) = e^{-\lambda}\frac{\lambda^j}{j!}$.

$$EVAL = 1 - \lambda \sum_{j=0}^{\infty} P_j(\lambda) \frac{2}{K+2j}$$

$$= 1 - \sum_{j=0}^{\infty} e^{-\lambda} \frac{\lambda^{j+1}}{j!} \frac{2}{K+2j} \frac{j+1}{j+1}$$

$$= 1 - \sum_{j=0}^{\infty} P_{j+1}(\lambda) \frac{2(j+1)}{K+2j}$$

Let $i = j+1$ and note that the following expression in the summation is zero when $i = 0$, therefore the index can be initially zero.

$$EVAL = 1 - \sum_{i=0}^{\infty} P_i(\lambda) \frac{2i}{K+2(i-1)}$$

$$= \sum_{i=0}^{\infty} P_i(\lambda)(1 - \frac{2i}{K+2i-2})$$

$$= \sum_{i=0}^{\infty} P_i(\lambda)(\frac{K+2i-2-2i}{K+2i-2})$$

$$= E(\frac{K-2}{K+2J-2})$$

Where J is distributed Poisson (λ) and Prob $(J=i) = P_i(\lambda)$.

Therefore,

$$RISK = K - 2b(K-2)E \frac{1}{2J+K-2} + b^2 E \frac{1}{2J+K-2}$$

Choose b to minimize Risk, let $S = E \frac{1}{2J+K-2}$.

$$\frac{\partial RISK}{\partial b} = -2S(K-2) + 2bS = 0$$

$$\frac{\partial^2 RISK}{\partial b^2} = 2S > 0$$

Solution is $b = K-2$ therefore

$$RISK = K - 2(K-2)^2 E \frac{1}{2J+K-2} + (K-2)^2 E \frac{1}{2J+K-2}$$

$$= K - (K-2)^2 E \frac{1}{2J+K-2}$$

Now $0 < E \frac{1}{2J+K-2} \le \frac{1}{K-2}$

Hence $2 \le RISK < K$.

G. EVALUATION OF $E\left(\frac{U}{W}\right)$

The goal of this section is to evaluate $E\left(\frac{U}{W}\right)$.

First, find the joint distribution of U and W.

$$f_{U,V}(u,v) = f_U(u) f_V(v)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-||\Theta||)^2} \frac{1}{2^{\frac{K-1}{2}}\Gamma\left(\frac{K-1}{2}\right)} v^{\frac{K-3}{2}} e^{-\frac{v}{2}} \qquad 0 < v < \infty \quad -\infty < u , \infty$$

Change of variables:

$$w = u^2 + v , \qquad v = w - u^2 , \qquad \frac{dv}{dw} = 1$$

$$0 < v < \infty \quad => \quad 0 < w - u^2 < \infty \quad => \quad u^2 < w < \infty$$

$$f_{U,W}(u,w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-||\Theta||)^2} \frac{1}{2^{\frac{K-1}{2}}\Gamma\left(\frac{K-1}{2}\right)} (w-u^2)^{\frac{K-3}{2}} e^{-\frac{1}{2}(w-u^2)}$$

$$-\infty < u < \infty \qquad u^2 < w < \infty$$

Since $-\frac{1}{2}(u-||\Theta||)^2 = -\frac{1}{2}(u^2-2u||\Theta||+||\Theta||^2)$ it follows that

$$f_{U,W}(u,w) = \frac{1}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\Gamma\left(\frac{K-1}{2}\right)} (w-u^2)^{\frac{K-3}{2}} e^{-\frac{1}{2}(w-2u||\Theta||+||\Theta||^2)}$$

Therefore,

$$E\left(\frac{U}{W}\right) = \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\Gamma\left(\frac{K-1}{2}\right)} \int_0^\infty e^{-\frac{w}{2}} \left[\int_{-\sqrt{w}}^{\sqrt{w}} \frac{u}{w}(w-u^2)^{\frac{K-3}{2}} e^{||\Theta|| u}\, du \right] dw$$

Where the range is $u^2 < w \;=>\; -\sqrt{w} < u < \sqrt{w}$.

Let $t = \frac{u}{\sqrt{w}}$, $dt = \frac{du}{\sqrt{w}}$, and $-1 < t < +1$. Continuing

$$E\left(\frac{U}{W}\right) = \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\Gamma\left(\frac{K-1}{2}\right)} \int_0^\infty e^{-\frac{w}{2}} \left[\int_{-1}^{1} \frac{t}{\sqrt{w}}\, w^{\frac{K-3}{2}} (1-t^2)^{\frac{K-3}{2}} e^{||\Theta|| t\sqrt{w}} \sqrt{w}\, dt \right] dw$$

$$= \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\Gamma\left(\frac{K-1}{2}\right)} \int_0^\infty e^{-\frac{w}{2}} \left[w^{\frac{K-3}{2}} \int_{-1}^{1} t(1-t^2)^{\frac{K-3}{2}} e^{||\Theta|| t\sqrt{w}}\, dt \right] dw$$

Now $e^{||\Theta|| t\sqrt{w}} = \sum_{i=0}^{\infty} \frac{||\Theta||^i}{i!} t^i w^{\frac{i}{2}}$.

$$E\left(\frac{U}{W}\right) = \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\, \Gamma(\frac{K-1}{2})} \int_0^\infty e^{-\frac{w}{2}} w^{\frac{K-3}{2}} \sum_{i=0}^\infty \frac{||\Theta||^i w^{\frac{i}{2}}}{i!} \left[\int_{-1}^{1} t^{i-1} (1-t^2)^{\frac{K-3}{2}} dt\right] dw$$

Let $i = 2j+1$.

$$E\left(\frac{U}{W}\right) = \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\, \Gamma(\frac{K-1}{2})} \int_0^\infty e^{-\frac{w}{2}} w^{\frac{K-3}{2}} \sum_{j=0}^\infty \frac{||\Theta||^{2j+1}}{(2j+1)!} w^{\frac{2j+1}{2}} \left[\int_{-1}^{1} t^{2(j+1)} (1-t^2)^{\frac{K-3}{2}} dt\right] dw$$

Now the inner integral is a $Beta\left[j+\frac{3}{2}, \frac{K-3}{2}+1\right] = Beta\left[j+\frac{3}{2}, \frac{K-1}{2}\right]$, and

$$Beta\left[j+\frac{3}{2}, \frac{K-1}{2}\right] = \frac{\Gamma(j+\frac{3}{2})\Gamma(\frac{K-1}{2})}{\Gamma(j+1+\frac{K}{2})}$$

$$E\left(\frac{U}{W}\right) = \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\, \Gamma(\frac{K-1}{2})} \Gamma\left(\frac{K-1}{2}\right) \sum_{j=0}^\infty \frac{||\Theta||^{2j+1}}{(2j+1)!} \frac{\Gamma(j+\frac{3}{2})}{\Gamma(j+1+\frac{K}{2})} \int_0^\infty w^{(\frac{K}{2}+j-1)} e^{-\frac{w}{2}} dw$$

The integral is $\dfrac{\Gamma(j+\frac{K}{2})}{(\frac{1}{2})^{j+\frac{K}{2}}} = 2^{j+\frac{K}{2}}\Gamma(j+\frac{K}{2})$. It follows that

$$E\left(\frac{U}{W}\right) = \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\, 2^{\frac{K-1}{2}}\, \Gamma(\frac{K-1}{2})} \Gamma\left(\frac{K-1}{2}\right) \sum_{j=0}^\infty 2^{j+\frac{K}{2}} \frac{||\Theta||^{2j+1}}{(2j+1)!} \frac{\Gamma(j+\frac{3}{2})}{j+\frac{K}{2}} .$$

Now,

$$\Gamma\left(j+\frac{3}{2}\right) = \left(j+\frac{1}{2}\right)\Gamma\left(j+\frac{1}{2}\right)$$

$$= \left(j+\frac{1}{2}\right)\left(j-\frac{1}{2}\right)\cdots\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)$$

$$\frac{1}{2^{j+1}}(2j+1)(2j-1)\cdots(3)(1)\,\Gamma\left(\frac{1}{2}\right)$$

Note further that

$$(2j+1)! = (2j+1)(2j)(2j-1)\cdots(3)(2)(1)$$

$$= (2j+1)(2j-1)\cdots(5)(3)(1)\ (2j)(2j-2)\cdots(6)(4)(2)$$

$$= \prod_{i=1}^{j}(2i+1)\ 2^j\ (j!)$$

$$E\left(\frac{U}{W}\right) = \frac{e^{-\frac{1}{2}||\Theta||^2}}{\sqrt{2\pi}\,2^{\frac{K-1}{2}}} \sum_{j=0}^{\infty} 2^{j+\frac{K}{2}} \frac{||\Theta||^{2j+1}}{\left[\prod_{i=1}^{j}(2i-1)\right]2^{j}\,(j!)} \frac{1}{2^{j+1}} \left[\prod_{i=1}^{j}(2i-1)\right]\Gamma\left(\frac{1}{2}\right)\frac{1}{j+\frac{K}{2}}$$

Since $\Gamma(\frac{1}{2})=\sqrt{\pi}$ and cancelling:

$$E\left(\frac{U}{W}\right) = e^{-\frac{1}{2}||\Theta||^2}\sum_{j=0}^{\infty}\frac{||\Theta||^{2j+1}}{2^{j+1}j!\,(j+\frac{K}{2})}$$

$$E\left(\frac{U}{W}\right) = ||\Theta|| \sum_{j=0}^{\infty} P_j(\lambda)\frac{1}{K-2j}$$

where $\lambda = \dfrac{||\Theta||^2}{2}$.

H. INVERSE SINE TRANSFORM

Method (1):

$$\Theta_1 = \sqrt{n}\ [2Sin^{-1}\sqrt{p}\ ]$$

Method (2):

$$\Theta_2 = \sqrt{n}\ [Sin^{-1}(2p-1) + \frac{\pi}{2}]$$

Method (2) was used because it was computationally faster for the computer to execute (2p-1) vice $\sqrt{p}$. Proof that both methods are the same is as follows. First step is to reduce the right hand side of each equation to the elements within the brackets.

$$Sin\left(\frac{\Theta_1}{\sqrt{n}}\right) = Sin\ [2Sin^{-1}\sqrt{p}\ ]$$

$$= 2\sqrt{p}\ \ Cos\ [Sin^{-1}\sqrt{p}\ ]$$

$$= 2\sqrt{p}\ \sqrt{1-p}$$

$$= 2\sqrt{p\,(1-p)}.$$

$$Sin\left(\frac{\Theta_2}{\sqrt{n}}\right) = Sin\ [Sin^{-1}(2p-1) - \frac{\pi}{2}]$$

$$= (2p-1)\ Cos\frac{\pi}{2} - Sin\frac{\pi}{2}\ Cos\ \ [Sin^{-1}(2p-1)]$$

$$= (2p-1)(0) - (1)\sqrt{1-(2p-1)^2}$$

$$= \sqrt{1-(4p^2-4p+1)}$$

$$= 2\sqrt{p\,(1-p)}.$$

Therefore $Sin\ \frac{\Theta_1}{\sqrt{n}} = Sin\ \frac{\Theta_2}{\sqrt{n}}$, implying $\Theta_1 = \Theta_2$. Q.E.D.

I. MONTHLY VERSES YEARLY RATES

This section provides a recommended method to convert yearly attrition rates to monthly rates.

Let r = yearly attrition rates.

Let r_m = monthly attrition rates.

Assume independence from month to month. Thus 12 consecutive survivor months implies a survivor year.

$$1 - r = (1 - r_m)^{12}$$

$$(1 - r)^{\frac{1}{12}} = 1 - r_m$$

$$r_m = 1 - (1 - r)^{\frac{1}{12}}$$

Thus, the above equation relates the yearly attrition rates to the monthly attrition rates.

APPENDIX D
PROGRAMS AND FUNCTIONS

This appendix contains sample listings of JCL, FORTRAN, AND APL programs utilized by the author throughout this research project.

A. ORIGINAL SUMMARY DATA PROGRAM

The original summary data from NPRDC is on a file in the Mass Storage at the NPGS Computer Center with a file name of "COUNTS". In order to access the data from CMS and transfer a copy of the data set "COUNTS" from Mass Storage to the MVS004 disk, submit the program "MSSCOUN JCL A1" to MVS. See Figure D.1.

The data set will nearly fill half a disk space of 4 cylinders, thus it is advisable to get a temporary work space while logged on in order to conserve A-disk space. The exec "GETTEMP EXEC A1" can be used by typing "GETTEMP 8 C". See Figure D.2. A temporary disk space of 8 cylinders will be useable with filemode "C". The files stored on this disk space will be accessible only while logged on. To view the files on this temporary disk space, type "FLIST * * C". This is similar to FLIST on the A-disk except it accesses the files on the temporary C-disk.

The "GETMVS" system exec is used to copy the summary data file to the temporary C-disk in order to access it from CMS. The exec will request the following identification information "S2209 COUNTS".

The summary data file has 16093 records with 53 characters per record. The number of records for each fiscal year is listed in Table XX.

```
      MSSCOUN   JCL  A
//COUNTS JOB (2209,5555),'MAJ D. D. TUCKER',CLASS=B
//*MAIN ORG=NPGVM1.2209P
//*  ********************** NOTE TO USERS ************************
//*   THIS WILL MOVE THE DATA SET COUNTS FROM MASS STORAGE TO THE
//*   MVS004 DISK.  FIRST CHANGE THE XXXX'S TO YOUR USER ID.
//*   CHANGE YYYY TO YOUR PROJECT NUMBER, THEN SUBMIT THE PROGRAM
//*   TO MVS.  WHEN THE SYSTEM RETURNS A COMPLETION MESSAGE, USE
//*   THE GETMVS SYSTEM EXEC TO COPY THE DATA SET TO YOUR A-DISK
//*   IF YOU WISH TO ACCESS IT FROM CMS.  BE ADVISED THAT THIS
//*   FILE IS HUGE AND WILL TAKE UP NEARLY 4 CYLINDERS.
//*
//*          USE DSN= S2209 COUNTS    IN GETMVS
//*  ***********************************************************
// EXEC PGM=IEBGENER
//SYSPRINT DD SYSOUT=A
//SYSIN DD DUMMY
//SYSUT1 DD DSN=MSS.F0968.COUNTS,DISP=OLD
//SYSUT2   DD DSN=S2209.COUNTS,VOL=SER=MVS004,UNIT=3350,
//         SPACE=(CYL,(2,2),RLSE),DISP=(NEW,KEEP)
/*
```

Figure D.1 MSSCOUN JCL A1.

B. DATA MANIPULATION PROGRAMS

A technique utilized by the author to reduce the dependency of manipulating such a large data file, involved creating a file of indices corresponding to pertinent data in the original summary data file. A set of indices was created for the inventories and losses identified for each fiscal year.

1. Inventory Indices

The set indices for the inventory can be created using the following procedures.

- Separate the summary data file by fiscal year data and create a new file for each fiscal year. For example, "COUNT77 DATA C" for fiscal year 1977 data on the temporary C-disk.

```
     GETTEMP   EXEC  A
&TRACE
*  GET &1 CYLINDERS OF TEMPORARY 199 DISK SPACE;
*  ACCESS IT AS MODE &2; DSK&2 AS DISK LABEL.
&IF .&1 = .? &GOTO -HELP
&IF &N NE 2 &GOTO -HELP
&IF &2 = A &GOTO -HELP
CP DEFINE T3350 AS 199 CYL &1
&IF &RC NE 0 &EXIT -1
&STACK YES
&LABEL = &CONCAT OF DSK &2
&IF &RC NE 0 &EXIT -2
&STACK &LABEL
FORMAT 199 &2
CLRSCRN
Q DISK
&EXIT
-HELP
     &TYPE ISSUE GETTEMP <A CYL> <FILEMODE>
     &TYPE WHERE CYL IS NUMBER OF CYLINDERS
     &TYPE FILEMODE = A IS NEVER ALLOWED
     &EXIT
&TRACE ON
```

Figure D.2   GETTEMP EXEC A1.

- Execute the NUMBER exec by typing "NUMBER <XX>" where XX is the desired fiscal year (such as 77 for fiscal year 1977 data). See Figure D.3. This executive will load the Fortran program "INV FORTRAN A1" which reads the data from the file "COUNT77 DATA C". See Figure D.4.
- The output from "INV FORTRAN A1" will be sent to a CMS file "INVXX ARRAY C", where XX is the desired fiscal year (same as above).

TABLE XX

RECORDS PER FISCAL YEAR

Fiscal Year	Number of Records
1977	2203
1978	2231
1979	2337
1980	2351
1981	2317
1982	2324
1983	2330

```
    FILE:  NUMBER  EXEC  A1

&TRACE OFF
*    TO EXECUTE THIS EXEC   TYPE  "NUMBER <YEAR>"
*     WHERE YEAR IS ONE OF 77-83.
GLOBAL TXTLIB VALTLIB VFORTLIB IMSLSP NONIMSL CMSLIB
GLOBAL LOADLIB VFLODLIB
FORTVS INV (LVL(77))
* FORTVS &1 (LVL(77))
CP TERM LINESIZE 133
&A = &CONCAT OF COUNT &1
&B = &CONCAT OF INV  &1
FILEDEF 08 DISK &A DATA C (RECFM F LRECL 53 PERM
FILEDEF 09 DISK &B ARRAY C (RECFM F LRECL 9 PERM
&TRACE ON
LOAD INV (START
* LOAD &1 (START
ERASE INV LISTING
ERASE INV TEXT
ERASE LOAD MAP
&EXIT
```

Figure D.3 NUMBER EXEC A1.

```
      FILE:  INV  FORTRAN  A1

C     PROGRAM INVENTORY
C     *****************************************************
C     *** REFORMATS DATA FROM 'COUNTS' FILE INTO A TWO    ***
C     *** DIMENTIONAL ARRAY OF INVENTORY FOR EACH YEAR.  ***
C     *** THE OUTPUT IS A SET OF INDICES FOR MATRIX       ***
C     *** FORMATION IN APL.                               ***
C     ***                                                 ***
C     ***               INDEX SUMMARY                     ***
C     ***  COL   CHAR          INFO                       ***
C     ***   1     1     YEAR (1,2,...,7)=(77,78,...,83)   ***
C     ***  2,3    2     MOS (00,01,...,39)                ***
C     ***   4     1     GRADE (0,1,...,9)                 ***
C     ***  5,6    2     LOS (0,1,...,30)                  ***
C     *** 7,8,9   3     INVENTORY NUMBER                  ***
C     ***                                                 ***
C     ***      USE "NUMBER  <YEAR>"     EXECUTIVE         ***
C     *****************************************************
C
C
      INTEGER GRADE, MOS, YEAR, LOS, INDEX, IN, TOT, IYR
C
C     THE DO LOOPS CORRESPOND TO THE NUMBER OF RECORDS
C         FOR YEARS 1977 TO 1983
C     THE FIRST ONE IS THE FULL DATA SET,
C         THE SECOND IS FOR TESTING
C
C     DO 200 I=1,16093
C     DO 200 I=1,50
C     DO 200 I=1,2203
C     DO 200 I=1,2231
C     DO 200 I=1,2337
C     DO 200 I=1,2351
C     DO 200 I=1,2317
C     DO 200 I=1,2324
      DO 200 I=1,2330
101   READ(8,60)YEAR,MOS,GRADE,LOS,INV,L1,L2,L3,L4,L5,L6,L7,L8
 60   FORMAT (4I2,9I5)
C
C     *** INITIALIZE THE INDEX  1=77,2=78,...,7=83
C
      IYR = YEAR - 76
      TOT = 100000000*IYR
C
      IF(INV.GT.0) CALL SUM(IN,TOT,MOS,GRADE,LOS,INV)
200   CONTINUE
C
      STOP
      END
C
C     THIS SUBROUTINE CREATES THE INDEX ARRAY
C         FOR AN INVENTORY DATA ELEMENT
C
      SUBROUTINE SUM(INDEX,I,J,K,L,NUM)
      INTEGER INDEX,I,J,K,L,NUM
        INDEX= I+(J*1000000)+(K*100000)+(L*1000)+NUM
      WRITE (9,500) INDEX
500   FORMAT (I9)
      RETURN
      END
```

Figure D.4 INV FORTRAN A1.

- The CMS file "INVXX ARRAY C", can be read into an APL workspace and assigned to a numeric variable by executing the user friendly "CMSAPL EXEC A1". See Figures D.5 and D.6. Note that "CMSAPL" has a provision to add a variable to an existing APL workspace or to create a new workspace each time "CMSAPL" is executed.

The file of indices created in this manner can be utilized to generate matrices in the APL workspace. The two dimensional array for each fiscal year's inventory is characterized by a length subject to the number of records in the summary data file for a particular fiscal year in which there is a nonzero inventory count. The columns of the records in the index file is described in Table XXI

2. Loss Indices

The set indices for the losses can be created in a manner similar to the indices for the inventory by using the following procedures.

- Separate the summary data file by fiscal year data and create a new file for each fiscal year. For example, "LOSS77 DATA C" for fiscal year 1977 data on the temporary C-disk.
- Execute the "LOSS EXEC A1" by typing "LOSS <XX>" where XX is the desired fiscal year (such as 77 for fiscal year 1977 data). See Figure D.7. This executive will load the Fortran program "LOSSES FORTRAN A1" which reads the data from the file "COUNT77 DATA C". See Figure D.8.
- The output from "LOSSES FORTRAN A1" will be sent to a CMS file "LOSSXX ARRAY C", where XX is the desired fiscal year (same as above).
- The CMS file "LOSSXX ARRAY C", can be read into an APL workspace using similar methods as the inventory file described above.

```
     CMSAPL   EXEC   A
&TRACE
&TYPE ENTER FILENAME, FILETYPE, MODE OF FILE TO BE READ INTO APL
&READ ARGS
&IF &N NE 3 &GOTO -TELL
STATE &1 &2 &3
&IF &RC NE 0 &GOTO -ERROR
&TYPE ENTER THE NAME OF THE APL VARIABLE THAT WILL STORE THE DATA
&READ VAR &A
&TYPE ENTER NAME OF APL WORKSPACE YOU DESIRE THE VARIABLE &A IN
&READ VAR &WKS
&TYPE ????  WHAT TYPE OF VARIABLE IS DESIRED?
&TYPE ENTER  N  IF YOU DESIRE A NUMERIC VARIABLE.
&TYPE ENTER  C  IF YOU DESIRE A CHARACTER VARIABLE.
&TYPE ENTER  Q  IF YOU DESIRE TO QUIT
&READ VAR &ASK
&IF &ASK EQ Q &EXIT
&TYPE YOUR INFORMATION WILL BE STORED IN APL WORKSPACE &WKS
&IF &ASK EQ N &GOTO -NUMS
CP TERMINAL APL ON
&STACK )LOAD 990 CMSIO
&STACK &A 'CMSREAD
&STACK &1
&STACK &2 &3
&STACK C
&STACK )WSID MYSPACE
&STACK )SAVE
&STACK )OFF HOLD
EXEC APL
&GOTO -SKIP
-NUMS CP TERMINAL APL ON
&STACK )LOAD 990 CMSIO
&STACK &A 'CMSREAD
&STACK &1
&STACK &2 &3
```

Figure D.5   CMSAPL EXEC A1.

```
&STACK N
&STACK )WSID MYSPACE
&STACK )SAVE
&STACK )OFF HOLD
EXEC APL
-SKIP &TYPE ????  IS THIS A NEW WORKSPACE?  Y OR N
&READ VAR &NEW
&IF &NEW EQ Y &GOTO -JUMP
CP TERMINAL APL ON
&STACK )LOAD &WKS
&STACK )PCOPY MYSPACE &A
&STACK )SAVE
&STACK )DROP MYSPACE
&STACK )OFF HOLD
EXEC APL
&EXIT 98
-JUMP CP TERMINAL APL ON
&STACK )WSID &WKS
&STACK )PCOPY MYSPACE &A
&STACK )SAVE
&STACK )DROP MYSPACE
&STACK )OFF HOLD
EXEC APL
&EXIT 99
-TELL &TYPE YOU HAVE ENTERED TOO MANY OR NOT ENOUGH ENTRIES
&TYPE FOR THE FILE THAT YOU WANT TO BE READ INTO APL.
&TYPE YOU NEED TO BEGIN AGAIN
&TYPE
&TYPE ENTER:     CMSAPL
&EXIT 100
-ERROR &TYPE &1 &2 &3 DOES NOT EXIS ON YOUR &CONCAT OF &3 -DISK
&TYPE CHECK YOUR FLIST AND THEN BEGIN AGAIN
&TYPE BY ENTERING          CMSAPL
&EXIT 101
```

Figure D.6   CMSAPL EXEC A1 (CONT.).

TABLE XXI

INVENTORY INDEX SUMMARY

Column	Characters	Information
1	1	YEAR (1=1977...7=1983)
2,3	2	MOS (00...39)
4	1	GRADE (0...9)
5,6	2	LOS (0...30)
7,8,9	3	INVENTORY NUMBER

```
    FILE: LOSS EXEC A1


&TRACE OFF
*    TO EXECUTE THIS EXEC   TYPE "LOSS <YEAR>"
*     WHERE YEAR IS ONE OF 77-83.
GLOBAL TXTLIB VALTLIB VFORTLIB IMSLSP NONIMSL CMSLIB
GLOBAL LOADLIB VFLODLIB
FORTVS LOSSES (LVL(77))
* FORTVS &1 (LVL(77))
CP TERM LINESIZE 133
&A = &CONCAT OF COUNT &1
&B = &CONCAT OF LOSS  &1
FILEDEF 08 DISK &A DATA C (RECFM F LRECL 53 PERM
FILEDEF 09 DISK &B ARRAY C (RECFM F LRECL 9 PERM
&TRACE ON
LOAD LOSSES (START
* LOAD &1 (START
ERASE LOSSES LISTING
ERASE LOSSES TEXT
ERASE LOAD MAP
&EXIT
```

Figure D.7   LOSS EXEC A1.

```
      PROGRAM LOSSES
C     ********************************************************
C     *** REFORMATS DATA FROM 'COUNTS' FILE INTO A TWO     ***
C     *** DIMENTIONAL ARRAY OF LOSSES  FOR EACH YEAR.      ***
C     *** THE OUTPUT IS A SET OF INDICES FOR MATRIX        ***
C     *** FORMATION IN APL.                                ***
C     ***                                                  ***
C     ***                 INDEX SUMMARY                    ***
C     ***  COL   CHAR             INFO                     ***
C     ***   1      1       YEAR (1,2,...,7)=(77,78,...,83) ***
C     ***  2,3     2       MOS (00,01,...,39)              ***
C     ***   4      1       GRADE (0,1,...,9)               ***
C     ***  5,6     2       LOS (0,1,...,30)                ***
C     *** 8,9,10   3       NUMBER OF LOSSES                ***
C     ***                                                  ***
C     ***      USE "LOSS <YEAR>"     EXECUTIVE             ***
C     ********************************************************
      INTEGER GRADE,MOS, YEAR, LOSS, LOS, IYR
C
C     THE DO LOOPS CORRESPOND TO THE NUMBER OF RECORDS FOR YEARS 77-83
C     THE FIRST ONE IS THE FULL DATA SET, THE SECOND IS FOR TESTING
C
C     DO 200 I=1,16093
C     DO 200 I=1,50
C     DO 200 I=1,2203
C     DO 200 I=1,2231
C     DO 200 I=1,2337
C     DO 200 I=1,2351
C     DO 200 I=1,2317
C     DO 200 I=1,2324
      DO 200 I=1,2330
C
C     ** L1 - L8 IS LOSS TYPE CODE 1 - 8
C
101   READ(8,60)YEAR,MOS,GRADE,LOS,INV,L1,L2,L3,L4,L5,L6,L7,L8
 60   FORMAT (4I2,9I5)
C
C     *** INITIALIZE THE INDEX  1=77,2=78,...,7=83
C
      IYR = YEAR - 76
      LOSS = 100000000*IYR
C
      IF(L1.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,1,L1)
      IF(L2.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,2,L2)
      IF(L3.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,3,L3)
      IF(L4.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,4,L4)
      IF(L5.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,5,L5)
      IF(L6.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,6,L6)
      IF(L7.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,7,L7)
      IF(L8.GT.0) CALL SUM(IN,LOSS,MOS,GRADE,LOS,8,L8)
200   CONTINUE
      STOP
      END
C
C     THIS SUBROUTINE CREATES THE INDEX ARRAY FOR LOSS DATA ELEMENT
C
      SUBROUTINE SUM(INDEX,I,J,K,L,M,NUM)
      INTEGER INDEX,I,J,K,L,M,NUM
         INDEX= I+(J*10000000)+(K*1000000)+(L*10000)+(M*1000)+NUM
      WRITE (9,500) INDEX
500   FORMAT (I10)
      RETURN
      END
```

Figure D.8 LOSSES FORTRAN A1.

The file of indices for the losses is similar to the inventory file. The two dimensional array for each fiscal year's loss is characterized by a length subject to the number of records in the summary data file for a particular fiscal year in which there is a nonzero loss count for each loss type. It should be noted that a record in the original

summary data may have a numerical count for more than one type of loss. The columns of the records in the index file is described in Table XXII.

TABLE XXII

LOSS INDEX SUMMARY

Column	Characters	Information
1	1	YEAR (1=1977...7=1983)
2,3	2	MOS (00...39)
4	1	GRADE (0...9)
5,6	2	LOS (0...30)
7	1	LOSS TYPE (1...8)
8,9,10	3	NUMBER OF LOSSES

C. FUNCTIONS

Numerous APL functions were utilized in this project for data manipulation and execution of calculations pertaining to the processes under evaluation. The following is a discussion of the most utilized functions and their purposes.

1. Inventory Functions

The workspace "INVTOT" was utilized to retain the inventory index matrices. The variables assigned to each fiscal year's inventory index matrix was labeled "INVXX", where "XX" is the applicable fiscal year (e.g., 77 for fiscal year 1977).

a. Function GETINV

The function to create the inventory matrix for each desired year is "GETINV". See Figure D.9. "GETINV" utilizes the function "INVMATX INVXX", which interprets the inventory index matrix "INVXX" and creates the inventory matrix for fiscal year "XX". See Figure D.10. The resultant matrix is "IXX" for fiscal year "XX". The function "INVMATX" could create a matrix of the following dimension 7x40x10x31 for 7 years, 40 MOS's, 10 grades, and 31 LOS's. However, due to limited workspace, the dimension of 40x31x10 for 40 MOS's, 31 LOS's, and 10 grades was commonly utilized.

```
      ∇ GETINV
[1]  ⍝ GET THE NUMBER OF YEARS OF INVENTORY DESIRED.
[2]  ⍝ COMMENT OUT YEARS NOT USED IMMEDIATELY.
[3]   I77←INVMATX INV77
[4]   I78←INVMATX INV78
[5]   I79←INVMATX INV79
[6]   I80←INVMATX INV80
[7]   I81←INVMATX INV81
[8]  ⍝ I82←INVMATX INV82
[9]  ⍝ I83←INVMATX INV83
[10]  ' SHAPE OF I77 IS '
[11]  ⎕←⍴I77
```

Figure D.9 APL Function GETINV.

b. Group Inventory Functions

There is a separate function to make a combined matrix for each of the MOS groups. Listed in Table XXIII are the groups, the function used to create the desired matrix, and the shape of the resultant matrix. Figure D.11 is a representation of the programing technique used for the

```
      ∇ Z←INVMATX X;A;B;C;D;E;F;I;J
[1]  ⍝'MATRIX INDEXING (YEAR,MOS,GRADE,LOS,INVENTORY) '
[2]   Z←(40 31 10)⍴0
[3]  ⍝ X IS AN ARRAY OF INDICES FOR A YEAR OF INVENTORY DATA.
[4]  ⍝ INDEX = 9 CHARACTERS PER ARRAY ELEMENT
[5]  ⍝ INDEX    NUMBER CHARACTERS           DESCRIPTION
[6]  ⍝   1              1            YEAR (1...7) FOR (77...83)
[7]  ⍝  2,3             2            MOS (00,01,...,39)
[8]  ⍝   4              1            GRADE (0,1,...,9)
[9]  ⍝  5,6             2            LOS (00,01,...,30)
[10] ⍝ 7,8,9            3            INVENTORY
[11] ⍝
[12]  I←⍴X
[13]  J←I÷9
[14] LOOP:→(J=0)/OUT
[15] ⍝  THIS INDEX IS NOT USED BECAUSE WITH THIS ADDED
[16] ⍝  DIMENSION, YOU'LL GET A FULL WORKSPACE.
[17]  A←⍎(1↑X)
[18] ⍝
[19]  B←1+(⍎(2↑X←(1↓X)))
[20]  C←1+(⍎(1↑X←(2↓X)))
[21]  D←1+(⍎(2↑X←(1↓X)))
[22]  E←⍎(3↑X←(2↓X))
[23]  Z[B;D;C]←E
[24]  X←(3↓X)
[25]  J←J−1
[26]  →LOOP
[27] OUT:'FINISHED -- SHAPE OF MATRIX IS   '
[28]  ⍴Z
```

Figure D.10 APL Function INVMATX.

Ground Combat MOS. The technique for the other groups was similar.

These functions call the function "GETMOS YY", which creates the central data inventory for the desired fiscal year for a particular MOS of "YY". In order to get a Combat Service Support MOS inventory matrix, the "GETMOS YY" function can be utilized separately to get the MOS "YY" matrix. The function "GETMOS" uses the global variables of "IXX" and "LXX", for the inventory matrix and loss matrix respectively, for fiscal year "XX". See Figure D.12.

```
     ∇ GETGC;K
[1]  ⍝ THIS GETS THE INVENTORY MATRIX FOR THE GROUND
[2]  ⍝ COMBAT OCCUPATION GROUP.
[3]  ⍝ SHAPE IS (MOS;YEARS;LOS;GRADE)
[4]  ⍝ ** CHECK THE 2ND INDEX FOR THE CORRECT
[5]  ⍝ ** NUMBER OF YEARS DESIRED.
[6]   GC←(3 4 31 10)⍴0
[7]  ⍝   THIS GETS THE MOS: INFANTRY
[8]   GC[1;;;]←GETMOS 3
[9]  ⍝   THIS GETS THE MOS: ARTILLERY
[10]  GC[2;;;]←GETMOS 5
[11] ⍝   THIS GETS THE MOS: TANKS AND AMPHIB
[12]  GC[3;;;]←GETMOS 10
```

Figure D.11 APL Function GETGC.

```
     ∇ Z←GETMOS X;J;Y
[1]  ⍝ GET THE CENTRAL DATA NUMBER OF INVENTORY
[2]  ⍝ FOR THE YEARS DESIRED FOR A PARTICULAR MOS.
[3]  ⍝ X= MOS DESIRED.  CHANGE THE INDEX FOR Z IF
[4]  ⍝ YEARS ARE NOT SEQUENTIAL.
[5]   J←X+1
[6]   Z←(4 31 10)⍴0
[7]   Z[1;;]←((Y←(I77[J;;]+I78[J;;])÷2)⌈(L77[J;;]))
[8]   Z[2;;]←((Y←(I78[J;;]+I79[J;;])÷2)⌈(L78[J;;]))
[9]   Z[3;;]←((Y←(I79[J;;]+I80[J;;])÷2)⌈(L79[J;;]))
[10]  Z[4;;]←((Y←(I80[J;;]+I81[J;;])÷2)⌈(L80[J;;]))
[11] ⍝Z[5;;]←((Y←(I81[J;;]+I82[J;;])÷2)⌈(L81[J;;]))
[12] ⍝Z[6;;]←((Y←(I82[J;;]+I83[J;;])÷2)⌈(L82[J;;]))
```

Figure D.12 APL Function GETMOS.

To create the inventory matrix used for the James-Stein function the following functions were used:

- GETNG - creates the "NG" matrix for Ground Combat. See Figure D.13 for an example of the program utilized to create the the Ground Combat inventory matrix. The

TABLE XXIII

INVENTORY FUNCTIONS

Occupation Group	Function Name	Shape
Aviators	GETAV	(Years,LOS,grade)
Combat Support	GETCS	(MOS,Years,LOS,grade)
Ground Combat	GETGC	(MOS,Years,LOS,grade)

Combat Support inventory matrix was created in a similar manner.

- GETNC - creates the "NC" matrix for Combat Support.

```
      ∇ GETNG
[1]   ⍝ THIS CREATES THE INVENTORY FOR GROUND
[2]   ⍝ COMBAT, YEARS, AND GRADE ELEMENTS.
[3]   ⍝ NG(X), X=0-8 FOR WO TO LT COL.
[4]   ⍝ NG0←(3 1 2)⍉GC[;;;1]
[5]   ⍝ NG1←(3 1 2)⍉GC[;;;2]
[6]   ⍝ NG2←(3 1 2)⍉GC[;;;3]
[7]   ⍝ NG3←(3 1 2)⍉GC[;;;4]
[8]   ⍝ NG4←(3 1 2)⍉GC[;;;5]
[9]     NG5←(3 1 2)⍉GC[;;;6]
[10]  ⍝ NG6←(3 1 2)⍉GC[;;;7]
[11]  ⍝ NG7←(3 1 2)⍉GC[;;;8]
[12]    NG8←(3 1 2)⍉GC[;;;9]
```

Figure D.13 APL Function GETNG.

2. Loss Functions

The workspace "MATRIX" was utilized to retain the loss index matrices. The variables assigned to each fiscal year's loss index matrix was labeled "LOSSXX", where "XX" is the applicable fiscal year (e.g., 77 for fiscal year 1977).

a. Function GETLOSS

The function to create the loss matrix for each desired year is "GETLOSS". See Figure D.14. "GETLOSS" utilizes the function "MATRIX LOSSXX", which interprets the loss index matrix "LOSSXX" and creates the loss matrix for fiscal year "XX". The resultant matrix is "LXX" for fiscal year "XX". The function "MATRIX" could create a matrix of the following dimension 7x40x10x31x8 for 7 years, 40 MOS's, 10 grades, 31 LOS's, and 8 loss types. However, due to limited workspace, the dimension of 40x31x10 for 40 MOS's, 31 LOS's, and 10 grades was commonly utilized when the losses were aggregated. Each loss type could be examined with the same dimensioned matrix. See Figure D.15.

```
    ∇ GETLOSS
[1]  ⍝ THIS GETS THE L(XX) MATRIX FOR YEARS
[2]  ⍝ XX=77,...,83.
[3]   L77←MATRIX LOSS77
[4]   L78←MATRIX LOSS78
[5]   L79←MATRIX LOSS79
[6]   L80←MATRIX LOSS80
[7]  ⍝L81←MATRIX LOSS81
[8]  ⍝L82←MATRIX LOSS82
[9]  ⍝L83←MATRIX LOSS83
```

Figure D.14 APL Function GETLOSS.

b. Group Loss Functions

There is a separate function to make a combined matrix for each of the MOS groups. Listed in Table XXIV are the groups, the function used to create the desired matrix, and the shape of the resultant matrix. See Figure D.16 for

```
1       ∇MATRIX[⎕]∇
        ∇ Z←MATRIX X;A;B;C;D;E;F;I;J
  [1]  ⍝ 'MATRIX INDEXING (YEAR,MOS,GRADE,LOS,LOSS TYPE,NUM OF LOSSES) '
  [2]  ⍝ USE FOLLOWING DIMENSION FOR Z IF SHAPE(MOS,GRADE,LOS,LOSSTYPE)
  [3]  ⍝ Z←(40 10 31 8)⍴0
  [4]  ⍝ USE THE FOLLOWING DIMENSION FOR Z IF AGGREGATING THE LOSSES
  [5]  ⍝ Z←(40 31 10)⍴0
  [6]  ⍝ USE THE FOLLOWING DIMENSION FOR Z IF SHAPE (YEAR,LOSSTYPE,GRADE)
  [7]  ⍝ Z←(1,8,10)⍴0
  [8]  ⍝ USE THE FOLLOWING DIMENSION FOR Z IF SHAPE (YEAR,LOSSTYPE,MOS)
  [9]  ⍝ Z←(1,8,40)⍴0
  [10] ⍝ USE THE FOLLOWING DIMENSION FOR Z IF SHAPE (YEAR,LOSSTYPE,LOS)
  [11]  Z←(1,8,31)⍴0
  [12] ⍝ X IS AN ARRAY OF INDICES FOR A YEAR OF LOSS DATA.
  [13] ⍝ INDEX = 10 CHARACTERS PER ARRAY ELEMENT
  [14] ⍝ INDEX   NUMBER CHARACTERS          DESCRIPTION
  [15] ⍝   1            1             YEAR (1,2,...,7) FOR (77,78,...,83)
  [16] ⍝  2,3           2             MOS (00,01,...,39)
  [17] ⍝   4            1             GRADE (0,1,...,9)
  [18] ⍝  5,6           2             LOS (00,01,...,30)
  [19] ⍝   7            1             LOSS TYPE (1,2,...,8)
  [20] ⍝ 8,9,10         3             NUMBER OF LOSSES
  [21] ⍝
  [22]  I←⍴X
  [23]  J←I÷10
  [24] LOOP:→(J=0)/OUT
  [25] ⍝  THIS INDEX IS NOT USED BECAUSE WITH THIS ADDED DIMENSION WS FULL
  [26] ⍝
  [27]  A←⍎(1↑X)
  [28] ⍝
  [29]  B←1+(⍎(2↑X←(1↓X)))
  [30]  C←1+(⍎(1↑X←(2↓X)))
  [31]  D←1+(⍎(2↑X←(1↓X)))
  [32]  E←⍎(1↑X←(2↓X))
  [33]  F←⍎(3↑X←(1↓X))
  [34] ⍝ USE THE FOLLOWING IF YOU WANT SEPARATE LOSSES
  [35] ⍝ NEED TO CHANGE THE DIMENSION OF Z ABOVE TO Z[40 10 31 8]
  [36] ⍝  Z[B;C,D;E]←F
  [37] ⍝ USE THE FOLLOWING IF YOU WANT THE LOSSES AGGREGATED
  [38] ⍝ Z[B;D;C]←Z[B;D;C]+F
  [39] ⍝ USE THE FOLLOWING IF YOU WANT SHAPE (YEAR,LOSSTYPE,GRADE)
  [40] ⍝ Z[1;E;C]←F
  [41] ⍝ USE THE FOLLOWING IF YOU WANT SHAPE (YEAR,LOSSTYPE,MOS)
  [42] ⍝ Z[1;E;B]←F
  [43] ⍝ USE THE FOLLOWING IF YOU WANT SHAPE (YEAR, LOSSTYPE,LOS)
  [44]  Z[1;E;D]←F
  [45]  X←(3↓X)
  [46]  J←J−1
  [47]  →LOOP
  [48] OUT:'FINISHED -- SHAPE OF MATRIX IS  '
```

Figure D.15 APL Function MATRIX.

the Ground Combat function. The other groups' functions were similar.

```
      ∇ GCLOSS
[1]  ⍝ THIS GETS THE LOSSES FOR GROUND COMBAT
[2]  ⍝ FOR 6 YEARS 77 - 82.
[3]  ⍝ SHAPE IS (MOS,YEAR,LOS,GRADE).
[4]  ⍝ CHANGE YEAR INDEX TO DESIRED NUMBER
[5]    GCL←(3 4 31 10)⍴0
[6]    GCL[1;1;;]←L77[4;;]
[7]    GCL[1;2;;]←L78[4;;]
[8]    GCL[1;3;;]←L79[4;;]
[9]    GCL[1;4;;]←L80[4;;]
[10]  ⍝ GCL[1;5;;]←L81[4;;]
[11]  ⍝ GCL[1;6;;]←L82[4;;]
[12]   GCL[2;1;;]←L77[6;;]
[13]   GCL[2;2;;]←L78[6;;]
[14]   GCL[2;3;;]←L79[6;;]
[15]   GCL[2;4;;]←L80[6;;]
[16] ⍝ GCL[2;5;;]←L81[6;;]
[17] ⍝ GCL[2;6;;]←L82[6;;]
[18]   GCL[3;1;;]←L77[11;;]
[19]   GCL[3;2;;]←L78[11;;]
[20]   GCL[3;3;;]←L79[11;;]
[21]   GCL[3;4;;]←L80[11;;]
[22] ⍝ GCL[3;5;;]←L81[11;;]
[23] ⍝ GCL[3;6;;]←L82[11;;]
```

Figure D.16 APL Function GCLOSS.

These functions use the global variables "LXX", for each "XX" fiscal year. In order to get a Combat Service Support MOS loss matrix, the "LXX" matrix was used directly to remove the desired combination of cells.

To create the loss matrix used for the James-Stein function the following functions were used:

- GETYC - creates the "YC" matrix for Combat Support.
- GETYG - creates the "YG" matrix for Ground Combat. See Figure D.17 for an example.

TABLE XXIV

LOSS FUNCTIONS

Occupation Group	Function Name	Shape
Combat Support	CSLOSS	(MOS,Years,LOS,grade)
Ground Combat	GCLOSS	(MOS,Years,LOS,grade)

```
      ∇ GETYG
[1]  ⍝ THIS CREATES THE LEAVERS FOR THE GROUND
[2]  ⍝ COMBAT, YEARS, AND GRADE ELEMENT.
[3]  ⍝ YG(X), X=0-8 FOR WO TO LTCOL.
[4]  ⍝ YG0←(3 1 2)⍉GCL[;;;1]
[5]  ⍝ YG1←(3 1 2)⍉GCL[;;;2]
[6]  ⍝ YG2←(3 1 2)⍉GCL[;;;3]
[7]  ⍝ YG3←(3 1 2)⍉GCL[;;;4]
[8]  ⍝ YG4←(3 1 2)⍉GCL[;;;5]
[9]    YG5←(3 1 2)⍉GCL[;;;6]
[10] ⍝ YG6←(3 1 2)⍉GCL[;;;7]
[11] ⍝ YG7←(3 1 2)⍉GCL[;;;8]
[12]   YG8←(3 1 2)⍉GCL[;;;9]
```

Figure D.17 APL Function GETYG.

3. James-Stein Functions

The following notation is required for the following functions.

- Let "i" stand for LOS, then i=0,...,30.
- Let "j" stand for MOS, (then values of j's depend on which MOS group is being analized).
- Let Dij = Incidence Matrix of nonstructural zeroes, which is the same for all years. Dij = 1, if cell is member of the feasible region; Dij = 0, otherwise.

- Let K = number of feasible cells, i.e., sum of all Dij.
- Let Yij = number of leavers in cell (i,j).
- Let t = 1,...,T; where T = number of years of data used to create the estimator.
- Let Nij = Inventory in cell (i,j) = Max ((N(t)+N(t+1))/2,Y(t)).

The following APL functions were utilized in the James-Stein model. See Figure D.18 for the APL listings of the following functions.

a. Function ARCSIN

Let ISM = Nij ARCSIN PMMij. The function "ARCSIN" returns the inverse sine transformation for use when the success probability is estimated directly (e.g., by MINMAX).

b. Function BINCONV

Let Pij = Jij BINCONV Nij(t). The function "BINCONV" inverts the inverse sine transformation.

c. Function BINPREP

Let ISij(t) = Yij(t) BINPREP Nij(t). The function "BINPREP" prepares the Freeman-Tukey version of the inverse sine transformation for binomial data.

d. Function JAMES

Let Jij = Dij JAMES ISij(t). The function "JAMES" returns a James-Stein estimator for the means of the means of the cells in the last two dimensions of the input matrix while being screened by the incidence matrix of nonstructural zeroes.

e. Function MINMAX

Let PMMij = Yij(t) MINMAX Nij(t). The function

```
     ∇ X←N ARCSIN P
[1]  ⍝ X IS THE INVERSE SINE TRANSFORMATION FOR USE WHEN THE
[2]  ⍝ SUCCESS PROB. P IS ESTIMATED DIRECTLY (E.G. BY MINIMAX)
[3]  ⍝ Y IS LOSSES; N IS INVENTORY
[4]   X←((0.5+N)*0.5)×−1○−1+2×P

     ∇ B←T BINCONV N;V0;V1
[1]  ⍝ INVERTS ARC SIN TRANSFORMATION
[2]  ⍝ T HAS RANK 2; N HAS RANK 3, THE 1st DIMENSION IS TIME.
[3]   B←0.5×1+1○T÷(0.5+(N←+⌿N)÷1↑⍴N)*0.5
[4]   V0←T<(−○÷2)×(N+0.5)*0.5
[5]   V1←T>(○÷2)×(N+0.5)*0.5
[6]   B←V1+B×V0⍲V1

     ∇BINPREP[⎕]∇
     ∇ X←Y BINPREP N;P
[1]  ⍝ PREPS THE FREEMAN−TUKEY VERSION OF THE
[2]  ⍝ ARC SIN TRANS FOR BINOMIAL DATA
[3]  ⍝ Y IS LOSSES; N IS INVENTORY
[4]   X←−1○−1+2×Y÷N+1
[5]   X←0.5×((0.5+N)*0.5)×X+−1○−1+2×(Y+1)÷N+1

     ∇ P←D JAMES Z;K;N;ZB;ZBB;S;M;C
[1]  ⍝ P IS A JAMES−STEIN ESTIMATOR FOR THE MEANS OF THE
[2]  ⍝ CELLS IN THE LAST TWO DIMENSIONS OF Z (SCREENED
[3]  ⍝ BY THE INCIDENCE MATRIX D OF NON STRUCTURAL ZEROS).
[4]   S←D SUMSQ Z
[5]   ZBB←(⍴ZB)⍴ZBB
[6]   C←1−(N×K−3)÷(2−K−K×N)×÷/S
[7]   P←D×ZBB+C×ZB−ZBB

     ∇ X←D SUMSQ Z;SSE
[1]  ⍝ X IS SSB,SSE FOR THE CELLS OF Z (WHICH HAS RANK 3)
[2]  ⍝ D (OF RANK 2) IS THE INCIDENCE MATRIX FOR THE LAST TWO
[3]  ⍝ DIMENSIONS OF Z. K IS THE NUMBER OF TREATMENTS.
[4]   K←+/,D
[5]   ZB←D×(+⌿Z)÷N←1↑⍴Z
[6]   ZBB←(+/,ZB)÷K
[7]   X←+/,(Z−(⍴Z)⍴ZBB)*2
[8]   X←(X−SSE),SSE←+/,(Z−(⍴Z)⍴ZB)*2
```

Figure D.18 APL Functions in James-Stein Algorithm.

"MINMAX" returns the MINMAX estimates for the binomial. The

function was utilized in the validation procedures to compare the James-Stein estimator to the MINMAX estimator. See Figure D.19 for the APL listing of this function.

```
     ∇ P←Y MINMAX N;YY
[1]  ⍝ RETURNS MINMAX EST'S FOR BINOMIAL
[2]  ⍝ INCLUDES REDUCTION OVER THE 1st RANK (TIME) OF THE
[3]  ⍝ INPUTS.  NN (GLOBAL) IS CREATED FOR USE IN ARC SINE.
[4]   YY←+⌿Y
[5]   NN←+⌿N
[6]   P←(0.5+(0≠YY)×YY÷NN*0.5)÷1+NN*0.5
```

Figure D.19 APL Function MINMAX.

4. Validation Functions

The following APL functions are an example of the functions utilized in the validation procedures specified in Chapter 5. Each MOS group (Aviation, Combat Support, Combat Service Support, and Ground Combat) was unique in its set of functions. However, the methodology was similar in manipulating the data.

a. Function BEFORE

The function "BEFORE" performed the preliminary work for the parameters for the James-Stein, Minimax, and Maximum Likelihood estimation processes. See Figure D.20.

b. Function GETGC81

The function "GETGC81" derives the Normal values for the James-Stein, Minimax, and Maximum Likelihood estimators. See Figure D.21.

```
        ∇ BEFORE;I5;I8
[1]   ⍝ DOES THE PRELIMINARY WORK FOR THE PARAMETERS
[2]   ⍝ FOR THE JAMES STEIN, MINMAX AND MAXLIKEHOOD
[3]   ⍝ FOR THE NORMALS OF THE GROUND COMBAT MOS.
[4]    I5←YG5 BINPREP NG5
[5]    I8←YG8 BINPREP NG8
[6]    JG5←DLT JAMES I5
[7]    JG8←DLC JAMES I8
[8]    PG5←JG5 BINCONV NG5
[9]    PG8←JG8 BINCONV NG8
[10]   PM5←YG5 MINMAX NG5
[11]   PM8←YG8 MINMAX NG8
[12]   IM5←(+/NG5) ARCSIN PM5
[13]   IM8←(+/NG8) ARCSIN PM8
[14]   IA5←(+/YG5) BINPREP(+/NG5)
[15]   IA8←(+/YG8) BINPREP(+/NG8)
[16]   PA5←IA5 BINCONV NG5
[17]   PA8←IA8 BINCONV NG8
```

Figure D.20 APL Function BEFORE.

c. Function AFTER

The function "AFTER" is an example of the consclioation of the execution of multiple years' validation functions. See Figure D.22.

d. Function GETFOM

The function "GETFOM" derives the figure of merit (i.e. mean squared plus variance) of a given array of values. See Figure D.23.

e. Function TEST

The function "TEST" was utilized in the additional comparison subtests mentioned in Chapter 5. The function conducted the two comparison tests for the projection capability of the James-Stein estimation process. See Figure D.24.

```
      ∇ GETGC81;CS81;Y;N5;N8;Y5;Y8;IG5;IG8
[1]   ⍝ GETS THE NORMALS FOR THE JAMES-STEIN, MINMAX,
[2]   ⍝ AND MAXLIKE FOR THE GROUND COMBAT MOS FOR THE
[3]   ⍝ PROJECTED YEAR 1981.  ALSO COMPUTES THE CHI
[4]   ⍝ SQUARE VALUES FOR VALIDATION PURPOSES.
[5]    I←0
[6]    N5←N81G5
[7]    N8←N81G8
[8]    Y5←Y81G5
[9]    Y8←Y81G8
[10]   IG5←(Y5 BINPREP N5)
[11]   IG8←(Y8 BINPREP N8)
[12]  NOR81G5←IG5-JG5
[13]  NOR81G8←IG8-JG8
[14]  DD←DLT
[15]  P←PG5
[16]  CS81G←(2 6)⍴0
[17]  CS81G[1; 1 2]←Y5 CHI N5
[18]  DD←DLC
[19]  P←PG8
[20]  CS81G[2; 1 2]←Y8 CHI N8
[21]  NMM81G5←IG5-IM5
[22]  NMM81G8←IG8-IM8
[23]  DD←DLT
[24]  P←PM5
[25]  CS81G[1; 3 4]←Y5 CHI N5
[26]  DD←DLC
[27]  P←PM8
[28]  CS81G[2; 3 4]←Y8 CHI N8
[29]  NA81G5←IG5-IA5
[30]  NA81G8←IG8-IA8
[31]  DD←DLT
[32]  P←PA5
[33]  CS81G[1; 5 6]←Y5 CHI N5
[34]  DD←DLC
[35]  P←PA8
[36]  CS81G[2; 5 6]←Y8 CHI N8
[37]  '  JAMES      DF      MINMAX      DF        MAXLIKE       DF'
[38]  ⎕←CS81G
```

Figure D.21 APL Function GETGC81.

```
      ∇ AFTER
[1]  ⍝ RUNS THE 81 TO 83 PROJECTION FUNCTIONS
[2]   'GROUND COMBAT 1981'
[3]   GETGC81
[4]   'GROUND COMBAT 1982'
[5]   GETGC82
[6]   'GROUND COMBAT 1983'
[7]   GETGC83
```

Figure D.22 APL Function AFTER.

```
      ∇ Z←GETFOM X;M;N;V
[1]  ⍝ GETS THE FIGURE OF MERIT (FOM) OF X
[2]  ⍝ FOM IS THE SQUARED MEAN PLUS THE VARIANCE
[3]   M←(+/,X)÷N←(⍴,X)
[4]   V←(+/,((X−M)*2))÷N−1
[5]   Z←V+(M*2)
[6]   'MEAN      = ',(⍕M)
[7]   'VARIANCE  = ',(⍕V)
[8]   'FOM       = ',(⍕Z)
```

Figure D.23 APL Function GETFOM.

```
∇ Z←Y TEST N;A;B;C;DD;IS;J;P;MLE;AC;SSEM;SEM;SSEP;SEP;DIFP;DIFM;NUM
  ⍝ Y IS LEAVERS, SHAPE (2 YEARS, 31 LOS, MOS'S)
  ⍝ D IS THE GLOBAL INCIDENT MATRIX FOR THE DESIRED GRADE
  ⍝ N IS CENTRAL INVENTORY, SHAPE (2 YEARS, 31 LOS, MOS'S)
  IS←Y BINPREP N
  J←D JAMES IS
  P←(J BINCONV N)×D
  MLE←((Y[1;;])÷(N[1;;]+(N[1;;]=0)))×D
  AC←((+⌿Y)÷((+⌿N)+((+⌿N)=0)))×D
  SSEM←+/,SEM←(AC−MLE)*2
  SSEP←+/,SEP←(AC−P)*2
  DIFP←+/,(SEP<SEM)
  DIFM←+/,(SEM<SEP)
  NUM←+/,D
  '1 THIS IS TEST 1, ALL THE CELLS IN THE FEASIBLE REGION'
  '  ARE ESTIMATED BY THE JAMES STEIN PROCESS'
  '  SEP = THE SQUARED DIFFERENCE OF THE JS PROJECTED AND ACTUAL'
  '  SEM = THE SQUARED DIFFERENCE OF THE MLE AND ACTUAL'
  '  SSE(MLE) = THE SUM OF THE SQUARED DIFFERENCES OF SEM'
  '  SSE(P) = THE SUM OF THE SQUARED DIFFERENCES OF SEP'
  ' '
  'FEASIBLE CELLS  SEP<SEM  SEM<SEP  SSE(MLE)  SSE(P)'
  A←'           '
  B←'        '
  C←'       '
  DD←'    '
  B,(⍕NUM),A,(⍕DIFP),B,(⍕DIFM),C,(⍕SSEM),DD,(⍕SSEP)
  ' '
  ' '
  SSEM←+/,SEM←(AC−(MLE×(AC≠0)))*2
  SSEP←+/,SEP←(AC−(P×(AC≠0)))*2
  DIFP←+/,(SEP<SEM)
  DIFM←+/,(SEM<SEP)
  NUM←+/,(AC≠0)
  ' THIS IS TEST 2, ONLY THE CELLS IN THE ACTUAL REGION >0'
  ' ARE REPRESENTED IN THE COMPARISON TEST '
  ' SEP2 = THE SQUARED DIFFERENCE OF THE JS PROJECTED AND ACTUAL'
  ' SEM2 = THE SQUARED DIFFERENCE OF THE MLE AND ACTUAL'
  ' SSE2(MLE) = THE SUM OF THE SQUARED DIFFERENCES OF SEM2'
  ' SSE2(P) = THE SUM OF THE SQUARED DIFFERENCES OF SEP2'
  ' '
  'FEASIBLE CELLS  SEP2<SEM2  SEM2<SEP2  SSE2(MLE)  SSE2(P)'
  A←'           '
  B←'        '
  C←'       '
  DD←'    '
  B,(⍕NUM),A,(⍕DIFP),B,(⍕DIFM),C,(⍕SSEM),DD,(⍕SSEP)
```

Figure D.24 APL Function TEST.

LIST OF REFERENCES

1. Bartholomew, D. J. and Forbes, A. F., Statistical Techniques for Manpower Planning, New York: Wiley, 1979.

2. Grinold, R. C. and Marshall, K. T., Manpower Planning Models, New York: North-Holand, 1977.

3. Law, A. and Kelton, W., Simulation Modeling and Analysis, New York: McGraw-Hill, 1982.

4. Larson, Harold J., Introduction to Probability Theory and Statistical Inference, New York: Wiley, 1969.

5. Stein, C., "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, v. 1, Berkeley: University of California Press, pp. 197-206, 1955.

6. James, W. and Stein, C., "Estimation with Quadratic Loss", Proceedings of the Fourth Berkeley Symposium, v. 1, University of California Press, 1961.

7. Efron, B. and Morris, C., "Data Analysis Using Stein's Estimator and its Generalizations," Journal of the American Statistical Association, v. 70, pp. 311-319, June 1975.

8. Efron, B. and Morris, C., "Stein's Paradox in Statistics," Scientific American, v. 236 (5) , pp. 119-127, 1977.

9. Berger, J. O., Statistical Decision Theory: Foundations, Concepts and Methods, New York: Springer-Verlag, 1980.

10. Efron, B. and Morris, C., "Limiting the Risk of Bayes and Empirical Bayes Estimators - Part II: The Empirical Bayes Case," Journal of the American Statistical Association, v. 67, pp. 130-139, March 1972.

11. Efron, B. and Morris, C., "Limiting the Risk of Bayes and Empirical Bayes Estimators - Part I: The Bayes Case," Journal of the American Statistical Association, v. 66, pp. 807-815, 1971.

BIBLIOGRAPHY

Berger, J. O., "Bayesian Robustness and the Stein Effect," Journal of the American Statistical Association, v. 77, pp. 358-368, June 1982.

Bishop, Y., Fienberg, S. and Holland, P., Discrete Multivariate Analysis: Theory and Practice, Cambridge: MIT Press, 1975.

Carter, G. and Rolph, J., "Empirical Bayes Methods Applied to Estimating Fire Alarm Probabilities," Journal of the American Statistical Association, v. 69, pp. 880-885, December 1974.

Casella, G., "An Introduction to Empirical Bayes Data Analysis," The American Statistician, v. 39, pp. 83-87, May 1985.

Freeman, M. F. and Tukey, J. W., "Transformations Related to the Angular and the Square Root", Annuals of Mathematical Statistics, v. 21, 1950.

Lehman, E. L. and Hodges, J. L. Jr, "Some Problems in Minimax Point Estimation," Annuals of Mathematical Statistics, v. 21, pp. 182-197, June 1950.

Morris, C., "Parametric Empirical Bayes Inference: Theory and Applications," Journal of the American Statistical Association, v. 78, pp. 47-65, March 1983.

Mosteller, F. and Tukey, J. W., Data Analysis and Regression: A Second Course in Statistics, Addison-Wesley, 1972.

Stein, C., "Estimation of the Mean of a Multivariate Normal Distribution," Annals of Statistics, v. 9, pp. 1135-1151, November 1981.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5100	2
3.	Professor Robert R. Read, Code 55Re Naval Postgraduate School Monterey, California 93943-5100	1
4.	Professor Paul R. Milch, Code 55Mh Naval Postgraduate School Monterey, California 93943-5100	1
5.	Marine Corps Representative, Code 0309 Naval Postgraduate School Monterey, California 93943-5100	1
6.	Commandant of the Marine Corps HQMC, Code MPP-30 Washington, D.C. 22314	1
7.	Commandant of the Marine Corps HQMC, Code MPI-10 Washington, D.C. 22314	1
8.	Commandant of the Marine Corps HQMC, Code MPI-20 Washington, D.C. 22314	1
9.	Commandant of the Marine Corps HQMC, Code MPI-40 Washington, D.C. 22314	1
10.	Commanding Officer Navy Personnel Research and Development Center San Diego, California 92152	1
11.	Major Dewey D. Tucker 9772 High Water Court Burke, Virginia 22015	1

END

DTIC

5-86